GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:21:39 ; Search time 29.6 Seconds
(without alignments)
106.594 Million cell updates/sec

Title: US-08-930-480A-7
Perfect score: 56
Sequence: 1 PKPSTPPGSS 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56	100.0	470	11	Q7TMK1	Q7tmk1 mus musculu
2	46	82.1	356	11	Q924W1	Q924w1 rattus norv
3	46	82.1	738	11	Q9Z213	Q9z213 rattus norv
4	46	82.1	1538	11	Q80U22	Q80u22 mus musculu
5	45	80.4	359	13	O73756	O73756 brachydanio
6	44	78.6	473	16	Q7W0Q8	Q7w0q8 bordetella
7	44	78.6	766	12	Q9DWF7	Q9dwf7 rat cytomeg
8	43	76.8	199	3	Q875A9	Q875a9 podospora a
9	43	76.8	209	5	O62173	O62173 caenorhabdi
10	43	76.8	372	11	Q91WK4	Q91wk4 mus musculu
11	43	76.8	587	16	Q93HI4	Q93hi4 streptomyce
12	43	76.8	984	4	Q8NH12	Q8nh12 homo sapien
13	42	75.0	153	10	Q8LDP2	Q8ldp2 arabidopsis
14	42	75.0	153	10	Q9LPH1	Q9lph1 arabidopsis
15	42	75.0	189	10	Q84JX4	Q84jx4 oryza sativ
16	42	75.0	237	16	Q8FDM6	Q8fdm6 escherichia
17	42	75.0	254	16	Q8FT08	Q8ft08 corynebacte
18	42	75.0	330	10	Q9LQT4	Q9lqt4 arabidopsis
19	42	75.0	418	16	O84876	O84876 chlamydia t
20	42	75.0	592	5	Q9VZ49	Q9vz49 drosophila
21	42	75.0	676	12	Q7T9D9	Q7t9d9 sudan ebola
22	42	75.0	955	4	Q96DN2	Q96dn2 homo sapien
23	42	75.0	1201	5	Q868S2	Q868s2 anopheles g
24	42	75.0	1638	5	Q7YYM5	Q7yym5 cryptospori
25	42	75.0	2838	5	Q8MP05	Q8mp05 tenebrio mo
26	41	73.2	96	17	Q972E4	Q972e4 sulfolobus
27	41	73.2	173	2	O85951	O85951 sphingomona
28	41	73.2	189	2	Q9L730	Q9l730 streptomyce
29	41	73.2	230	16	Q7U7W0	Q7u7w0 synechococc
30	41	73.2	330	4	Q96D28	Q96d28 homo sapien
31	41	73.2	339	5	Q9U156	Q9u156 leishmania
32	41	73.2	458	10	Q9FRJ1	Q9frj1 oryza sativ
33	41	73.2	458	10	Q7XCM0	Q7xcm0 oryza sativ
34	41	73.2	476	5	Q9VTU3	Q9vtu3 drosophila
35	41	73.2	582	16	Q82EA4	Q82ea4 streptomyce
36	41	73.2	737	13	Q90422	Q90422 brachydanio
37	41	73.2	750	13	Q9W633	Q9w633 cyprinus ca
38	41	73.2	762	13	Q9YIC6	Q9yic6 cyprinus ca
39	41	73.2	922	11	Q8BZX2	Q8bzx2 mus musculu
40	41	73.2	1984	12	Q9YQ45	Q9yq45 viral hemor
41	40	71.4	73	15	Q902F7	Q902f7 simian t-ly
42	40	71.4	73	15	Q8UU17	Q8uu17 simian t-ly
43	40	71.4	86	15	O38316	O38316 human immun
44	40	71.4	100	16	Q9FCM1	Q9fcm1 streptomyce
45	40	71.4	122	10	O24523	O24523 oryza sativ

ALIGNMENTS

RESULT 1
Q7TMK1
ID Q7TMK1 PRELIMINARY; PRT; 470 AA.
AC Q7TMK1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Breast tumor;

```
RA   Strausberg R.;
RL   Submitted (AUG-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC055910; AAH55910.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   470 AA;  51727 MW;  6D90E4DF896BB090 CRC64;

  Query Match             100.0%;  Score 56;  DB 11;  Length 470;
  Best Local Similarity   100.0%;  Pred. No. 0.57;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              ||||||||||
Db        242 PKPSTPPGSS 251


RESULT 2
Q924W1
ID   Q924W1      PRELIMINARY;      PRT;   356 AA.
AC   Q924W1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ALEX protein.
GN   ALEX.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95089824; PubMed=7997272;
RA   Kehlenbach R.H., Matthey J., Huttner W.B.;
RT   "XLas is a new type of G protein.";
RL   Nature 372:804-809(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Klemke M., Kehlenbach R.H., Huttner W.B.;
RT   "Two overlapping reading frames in a single exon encode interacting
RT   proteins - a novel way of gene usage.";
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X84047; CAC39212.1; -.
SQ   SEQUENCE   356 AA;  37970 MW;  9849ABD0AE524A3D CRC64;

  Query Match             82.1%;  Score 46;  DB 11;  Length 356;
  Best Local Similarity   80.0%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              |||| ||| |
Db         64 PKPSLPPGQS 73


RESULT 3
Q9Z213
ID   Q9Z213      PRELIMINARY;      PRT;   738 AA.
AC   Q9Z213;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RA   Wang Y.Z., Kehlenbach R.H., Huttner W.B.;
RT   "The XL-domain of rat XLas is encoded by a single exon.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF093569; AAD03033.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   738 AA;  80340 MW;  51EA2B3A7D9D018A CRC64;

  Query Match             82.1%;  Score 46;  DB 11;  Length 738;
  Best Local Similarity   80.0%;  Pred. No. 35;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              |||| ||| |
Db        446 PKPSLPPGQS 455


RESULT 4
Q80U22
ID   Q80U22      PRELIMINARY;      PRT;  1538 AA.
AC   Q80U22;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MKIAA0375 protein (Fragment).
GN   MKIAA0375.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22579291; PubMed=12693553;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA   Nakajima D., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 10:35-48(2003).
DR   EMBL; AK122263; BAC65545.1; -.
DR   InterPro; IPR004012; Run.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF02759; RUN; 1.
DR   Pfam; PF00018; SH3; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00593; RUN; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50826; RUN; 1.
DR   PROSITE; PS50002; SH3; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1538 AA;  163862 MW;  FBF93F32E3CD8EEE CRC64;

  Query Match             82.1%;  Score 46;  DB 11;  Length 1538;
  Best Local Similarity   80.0%;  Pred. No. 73;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              | ||:|||||
Db       1527 PTPSSPPGSS 1536


RESULT 5
O73756
ID   O73756      PRELIMINARY;      PRT;   359 AA.
AC   O73756;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Complement component BfB (Fragment).
GN   BFB OR BFB.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KOC;
```

```
RX   MEDLINE=99089896; PubMed=9874500;
RA   Gongora R., Figueroa F., Klein J.;
RT   "Independent duplications of Bf and C3 complement genes in the
RT   zebrafish.";
RL   Scand. J. Immunol. 48:651-658(1998).
DR   EMBL; AF047412; AAC05096.1; -.
DR   ZFIN; ZDB-GENE-990415-34; bfb.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00084; sushi; 2.
DR   Pfam; PF00092; vwa; 1.
DR   SMART; SM00032; CCP; 2.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50234; VWFA; 1.
FT   NON_TER       1      1
FT   NON_TER     359    359
SQ   SEQUENCE   359 AA;  40877 MW;  48770B63F310E10D CRC64;

  Query Match             80.4%;  Score 45;  DB 13;  Length 359;
  Best Local Similarity   80.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              | || |||||
Db         60 PDPSVPPGSS 69


RESULT 6
Q7W0Q8
ID   Q7W0Q8      PRELIMINARY;      PRT;   473 AA.
AC   Q7W0Q8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   BP0030.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640411; CAE40409.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   473 AA;  50610 MW;  EE84AED62A730A9F CRC64;

  Query Match             78.6%;  Score 44;  DB 16;  Length 473;
  Best Local Similarity   77.8%;  Pred. No. 46;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | ||||||:
Db          5 PSPSTPPGN 13


RESULT 7
Q9DWF7
ID   Q9DWF7      PRELIMINARY;      PRT;   766 AA.
AC   Q9DWF7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   PR34.
GN   R34.
OS   Rat cytomegalovirus (strain Maastricht).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Muromegalovirus.
OX   NCBI_TaxID=79700;
RN   [1]
RP   SEQUENCE OF 1-6 FROM N.A.
RC   STRAIN=Maastricht;
RX   MEDLINE=98139136; PubMed=9499096;
RA   Beisser P.S., Vink C., Van Dam J.G., Grauls G., Vanherle S.J.,
RA   Bruggeman C.A.;
RT   "The R33 G protein-coupled receptor gene of rat cytomegalovirus plays
RT   an essential role in the pathogenesis of viral infection.";
RL   J. Virol. 72:2352-2363(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Maastricht;
RX   MEDLINE=20366325; PubMed=10906222;
RA   Vink C., Beuken E., Bruggeman C.A.;
RT   "Complete DNA sequence of the rat cytomegalovirus genome.";
RL   J. Virol. 74:7656-7665(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Maastricht;
RX   MEDLINE=20473137; PubMed=11018281;
RA   Gruijthuijsen Y.K., Beuken E., Bruggeman C.A., Vink C.;
RT   "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
RT   spliced transcript.";
RL   Virus Res. 69:119-130(2000).
DR   EMBL; AF232689; AAF99132.1; -.
SQ   SEQUENCE   766 AA;  84603 MW;  101EC58097524704 CRC64;

  Query Match             78.6%;  Score 44;  DB 12;  Length 766;
  Best Local Similarity   70.0%;  Pred. No. 75;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              |:|| ||||:
Db        154 PRPSRPPGSA 163


RESULT 8
Q875A9
ID   Q875A9      PRELIMINARY;      PRT;   199 AA.
AC   Q875A9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to snoRNP protein gar1 of Schizosaccharomyces pombe.
OS   Podospora anserina.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX   NCBI_TaxID=5145;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Genoscope;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX088700; CAD60706.1; -.
DR   InterPro; IPR007504; Gar1.
DR   Pfam; PF04410; Gar1; 1.
SQ   SEQUENCE   199 AA;  20050 MW;  7E7E7E5779A2B05F CRC64;

  Query Match             76.8%;  Score 43;  DB 3;  Length 199;
  Best Local Similarity   70.0%;  Pred. No. 28;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              |||  |||:|
```

```
Db        110 PKPKPPPGAS 119


RESULT 9
O62173
ID   O62173      PRELIMINARY;      PRT;   209 AA.
AC   O62173;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   F15D3.6 protein.
GN   F15D3.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   White S.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81063; CAB02955.1; -.
DR   PIR; T20975; T20975.
DR   WormPep; F15D3.6; CE15853.
DR   InterPro; IPR006797; MSF1.
DR   Pfam; PF04707; MSF1; 1.
DR   PROSITE; PS50904; PRELI_MSF1; 1.
SQ   SEQUENCE   209 AA;  23667 MW;  346AB71D4BBD39C1 CRC64;

  Query Match            76.8%;  Score 43;  DB 5;  Length 209;
  Best Local Similarity  80.0%;  Pred. No. 29;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              |||||||  |
Db        199 PKPSTPPPPS 208


RESULT 10
Q91WK4
ID   Q91WK4      PRELIMINARY;      PRT;   372 AA.
AC   Q91WK4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Microtubule-associated protein tau.
GN   MAPT.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye, and Retina;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC014748; AAH14748.1; -.
DR   MGD; MGI:97180; Mapt.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   InterPro; IPR002955; Tau_protein.
DR   InterPro; IPR001084; Tubulin_Tau.
DR   Pfam; PF00418; tubulin-binding; 4.
DR   PRINTS; PR01261; TAUPROTEIN.
DR   PROSITE; PS00229; TAU_MAP; 4.
SQ   SEQUENCE   372 AA;  38861 MW;  B027745D23BC62A2 CRC64;

  Query Match            76.8%;  Score 43;  DB 11;  Length 372;
  Best Local Similarity  77.8%;  Pred. No. 52;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | | |||||
Db        108 PSPKTPPGS 116


RESULT 11
Q93HI4
ID   Q93HI4      PRELIMINARY;      PRT;   587 AA.
AC   Q93HI4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ABC transport protein (Putative ABC transporter ATP-binding
DE   protein).
GN   SAV7357.
OS   Streptomyces avermitilis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=22608306; PubMed=12692562;
RA   Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA   Sakaki Y., Hattori M., Omura S.;
RT   "Complete genome sequence and comparative analysis of the industrial
RT   microorganism Streptomyces avermitilis.";
RL   Nat. Biotechnol. 21:526-531(2003).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR   EMBL; AB070941; BAB69203.1; -.
DR   EMBL; AP005050; BAC75068.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001140; ABC_TM_transpt.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00664; ABC_membrane; 1.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW   ATP-binding; Transport; Complete proteome.
SQ   SEQUENCE   587 AA;  61099 MW;  18FB89332A4EC0C1 CRC64;

  Query Match            76.8%;  Score 43;  DB 16;  Length 587;
  Best Local Similarity  77.8%;  Pred. No. 83;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | |:|||||
Db          5 PTPTTPPGS 13


RESULT 12
Q8NH12
ID   Q8NH12      PRELIMINARY;      PRT;   984 AA.
```

```
AC   Q8NH12;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Seven transmembrane helix receptor.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA   Tsutsumi S., Aburatani H., Asai K., Akiyama Y.;
RT   "Genome-wide discovery and analysis of human seven transmembrane helix
RT   receptor genes.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB065601; BAC05829.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0007218; P:neuropeptide signaling pathway; IEA.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   Receptor; Transmembrane.
SQ   SEQUENCE   984 AA;  104440 MW;  01A78846EA48B155 CRC64;

  Query Match             76.8%;  Score 43;  DB 4;  Length 984;
  Best Local Similarity   70.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              | | :|||||
Db        635 PSPGSPPGSS 644


RESULT 13
Q8LDP2
ID   Q8LDP2      PRELIMINARY;      PRT;   153 AA.
AC   Q8LDP2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY085879; AAM63092.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   153 AA;  16475 MW;  730128E4C948C067 CRC64;

  Query Match             75.0%;  Score 42;  DB 10;  Length 153;
  Best Local Similarity   77.8%;  Pred. No. 31;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              |:|| ||||
Db         43 PQPSPPPGS 51


RESULT 14
Q9LPH1
ID   Q9LPH1      PRELIMINARY;      PRT;   153 AA.
AC   Q9LPH1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   T3F20.13 protein (Hypothetical protein) (At1g53560).
GN   T3F20.13 OR F22G10.10 OR AT1G53560.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lee J.M., Vaysberg M., Sakano H., Lenz C., Liu S.X., Pham P.,
RA   Toriumi M., Yu G., Chin C., Chiou J., Choi E., Chung M., Gonzalez A.,
RA   Howng B., Liu A., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA   Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA   Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
RA   Ecker J.R., Federspiel N.A., Theologis A.;
RT   "The sequence of BAC T3F20 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA   Dale J.M., Hayashizaki Y., Hsuan V.W., Ishida J., Jones T., Kamiya A.,
RA   Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA   Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis ORF clones.";
```

```
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC018748; AAF78434.1; -.
DR   EMBL; AC024260; AAG51969.1; -.
DR   EMBL; AY099799; AAM20650.1; -.
DR   EMBL; BT006546; AAP21354.1; -.
DR   PIR; F96575; F96575.
KW   Hypothetical protein.
SQ   SEQUENCE   153 AA;  16479 MW;  321129E4D528DD0B CRC64;

  Query Match             75.0%;  Score 42;  DB 10;  Length 153;
  Best Local Similarity   77.8%;  Pred. No. 31;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              |:|| ||||
Db         43 PQPSPPPGS 51


RESULT 15
Q84JX4
ID   Q84JX4      PRELIMINARY;      PRT;   189 AA.
AC   Q84JX4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   OJ1372_D12.32 protein (OJ1372_D12.41 protein) (OJ1372_D12.50
DE   protein).
GN   OJ1372_D12.32 OR OJ1372_D12.41 OR OJ1372_D12.50.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
RT   clone:OJ1372_D12.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003827; BAC57668.1; -.
DR   EMBL; AP003827; BAC57672.1; -.
DR   EMBL; AP003827; BAC57676.1; -.
SQ   SEQUENCE   189 AA;  20940 MW;  544F167C4AC7C49B CRC64;

  Query Match             75.0%;  Score 42;  DB 10;  Length 189;
  Best Local Similarity   70.0%;  Pred. No. 38;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              | |:|||||:
Db          5 PWPTTPPGSA 14


Search completed: April 20, 2004, 10:27:03
Job time : 30.6 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:21:04 ; Search time 8.4 Seconds
(without alignments)
61.988 Million cell updates/sec

Title: US-08-930-480A-7
Perfect score: 56
Sequence: 1 PKPSTPPGSS 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56	100.0	329	1	GC3_MOUSE	P22436 mus musculu
2	56	100.0	398	1	GC3M_MOUSE	P03987 mus musculu
3	43	76.8	732	1	TAU_MOUSE	P10637 mus musculu
4	43	76.8	751	1	TAU_RAT	P19332 rattus norv
5	42	75.0	237	1	YGHS_ECOLI	Q46843 escherichia
6	41	73.2	441	1	GUN2_THEFU	P26222 thermomonos
7	40	71.4	386	1	SYT5_HUMAN	O00445 homo sapien
8	40	71.4	476	1	TRZA_RHOCO	Q52725 rhodococcus
9	40	71.4	534	1	APG_ARATH	P40602 arabidopsis
10	40	71.4	579	1	COE2_BRARE	O93375 brachydanio
11	40	71.4	592	1	FZD9_MOUSE	Q9r216 mus musculu
12	40	71.4	601	1	3BP1_MOUSE	P55194 mus musculu
13	40	71.4	1036	1	ACK1_HUMAN	Q07912 homo sapien
14	40	71.4	1055	1	ACK1_MOUSE	O54967 mus musculu
15	40	71.4	1167	1	SOR1_MOUSE	Q9jlc4 mus musculu
16	39.5	70.5	364	1	IE68_PRVKA	P24827 pseudorabie
17	39	69.6	181	1	RRP3_HORVU	O48609 hordeum vul
18	39	69.6	307	1	CC36_CAEEL	P34803 caenorhabdi
19	39	69.6	386	1	SYT5_MOUSE	Q9r0n5 mus musculu
20	39	69.6	386	1	SYT5_RAT	P47861 rattus norv
21	39	69.6	489	1	OCLN_POTTR	Q28793 potorous tr
22	39	69.6	532	1	GRB7_HUMAN	Q14451 homo sapien
23	39	69.6	536	1	GAG_FSVMD	P03340 feline sarc
24	39	69.6	778	1	GELS_CHICK	O93510 gallus gall
25	39	69.6	790	1	KIF9_MOUSE	Q9wv04 mus musculu
26	39	69.6	860	1	AREA_PENRO	O13508 penicillium
27	39	69.6	865	1	NRFA_PENUR	Q92269 penicillium
28	39	69.6	1167	1	SOR1_HUMAN	Q8wy21 homo sapien
29	39	69.6	3149	1	TEGU_EBV	P03186 epstein-bar
30	38	67.9	126	1	DOCR_HUMAN	O75956 homo sapien
31	38	67.9	240	1	PRA_MYCTU	O53426 mycobacteri
32	38	67.9	349	1	ATF4_MOUSE	Q06507 mus musculu
33	38	67.9	353	1	ALC1_GORGO	P20758 gorilla gor
34	38	67.9	353	1	ALC1_HUMAN	P01876 homo sapien
35	38	67.9	391	1	P53_RABIT	Q95330 oryctolagus
36	38	67.9	546	1	COAT_SOUV3	Q04542 southampton
37	38	67.9	649	1	TFE2_MESAU	P98180 mesocricetu
38	38	67.9	649	1	TFE2_RAT	P21677 rattus norv
39	38	67.9	886	1	POL_FOAMV	P14350 human spuma
40	38	67.9	892	1	HIC1_MOUSE	Q9r1y5 mus musculu
41	38	67.9	1043	1	Y483_CHLPN	Q9z868 chlamydia p
42	38	67.9	1210	1	AF4_HUMAN	P51825 homo sapien
43	38	67.9	1415	1	ICP4_HSVMG	Q02362 marek's dis
44	38	67.9	1822	1	ZAP3_HUMAN	P49750 homo sapien
45	37	66.1	242	1	GUN5_TRIRE	P43317 trichoderma

ALIGNMENTS

RESULT 1
GC3_MOUSE
ID GC3_MOUSE STANDARD; PRT; 329 AA.
AC P22436;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig gamma-3 chain C region, secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85027161; PubMed=6092053;
RA Wels J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,
RA Tucker P.W., Blattner F.R.;
RT "Structural analysis of the murine IgG3 constant region gene.";
RL EMBO J. 3:2041-2046(1984).

DR EMBL; J00451; -; NOT_ANNOTATED_CDS.
DR PIR; B02156; G3MSC.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; ig; 3.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Transmembrane; Alternative splicing.
FT NON_TER 1 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 113 HINGE.
FT DOMAIN 114 223 CH2.
FT DOMAIN 224 327 CH3.
SQ SEQUENCE 329 AA; 36228 MW; F45827174182BAD6 CRC64;

Query Match 100.0%; Score 56; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 PKPSTPPGSS 10
              ||||||||||
Db        101 PKPSTPPGSS 110
```

```
RESULT 2
GC3M_MOUSE
ID   GC3M_MOUSE      STANDARD;      PRT;   398 AA.
AC   P03987;
DT   23-OCT-1986 (Rel. 02, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig gamma-3 chain C region, membrane-bound form.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85027161; PubMed=6092053;
RA   Wels J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,
RA   Tucker P.W., Blattner F.R.;
RT   "Structural analysis of the murine IgG3 constant region gene.";
RL   EMBO J. 3:2041-2046(1984).
RN   [2]
RP   SEQUENCE OF 328-398 FROM N.A.
RX   MEDLINE=84041483; PubMed=6314258;
RA   Komaromy M., Clayton L., Rogers J., Robertson S., Kettman J.,
RA   Wall R.;
RT   "The structure of the mouse immunoglobulin in gamma 3 membrane gene
RT   segment.";
RL   Nucleic Acids Res. 11:6775-6785(1983).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J00451; AAB59655.1; -.
DR   EMBL; V01526; CAA24767.1; ALT_SEQ.
DR   PIR; A02156; G3MSM.
DR   HSSP; P01857; 1FC1.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003597; Ig_c1.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00407; IGc1; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW   Transmembrane; Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN        1     97       CH1.
FT   DOMAIN       98    113       HINGE.
FT   DOMAIN      114    223       CH2.
FT   DOMAIN      224    327       CH3.
FT   TRANSMEM    346    362       POTENTIAL.
FT   DOMAIN      363    398       CYTOPLASMIC (POTENTIAL).
FT   CONFLICT    333    333       E -> G (IN REF. 2).
FT   CONFLICT    342    342       E -> Q (IN REF. 2).
FT   CONFLICT    388    388       P -> F (IN REF. 2).
SQ   SEQUENCE   398 AA;  43929 MW;  CF7F264B50A41B95 CRC64;

  Query Match             100.0%;  Score 56;  DB 1;  Length 398;
  Best Local Similarity   100.0%;  Pred. No. 0.53;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              ||||||||||
Db        101 PKPSTPPGSS 110


RESULT 3
TAU_MOUSE
ID   TAU_MOUSE       STANDARD;      PRT;   732 AA.
AC   P10637; P10638; Q60684; Q60685; Q60686; Q62286;
DT   01-JUL-1989 (Rel. 11, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Microtubule-associated protein tau (Neurofibrillary tangle protein)
DE   (Paired helical filament-tau) (PHF-tau).
GN   MAPT OR MTAPT OR TAU.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM PNS-TAU).
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=92262443; PubMed=1374898;
RA   Couchie D., Mavilia C., Georgieff I.S., Liem R.K.H., Shelanski M.L.,
RA   Nunez J.;
RT   "Primary structure of high molecular weight tau present in the
RT   peripheral nervous system.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:4378-4381(1992).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS TAU-A; TAU-D AND TAU-E).
RC   STRAIN=Him OF1; TISSUE=Brain, Kidney, and Liver;
RX   MEDLINE=95012085; PubMed=7927211;
RA   Kenner L., el-Shabrawi Y., Hutter H., Forstner M., Zatloukal K.,
RA   Hoefler G., Preisegger K.-H., Kurzbauer R., Denk H.;
RT   "Expression of three- and four-repeat tau isoforms in mouse liver.";
RL   Hepatology 20:1086-1089(1994).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS TAU-B AND TAU-C).
RC   TISSUE=Brain;
RX   MEDLINE=88099510; PubMed=3122323;
RA   Lee G., Cowan N.J., Kirschner M.;
RT   "The primary structure and heterogeneity of tau protein from mouse
RT   brain.";
RL   Science 239:285-288(1988).
RN   [4]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORM B).
RC   STRAIN=ICR; TISSUE=Brain;
RX   MEDLINE=95182802; PubMed=7877441;
RA   Sawa A., Oyama F., Matsushita M., Ihara Y.;
RT   "Molecular diversity at the carboxyl terminus of human and rat tau.";
RL   Brain Res. Mol. Brain Res. 27:111-117(1994).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=94005827; PubMed=8402267;
RA   Couchie D., Gache Y., Mavilia C., Guilleminot J., Bridoux A.-M.,
RA   Nivez M.-P., Nunez J.;
RT   "High molecular weight tau proteins and acquisition of neuronal
RT   polarity in peripheral nervous system.";
RL   C. R. Acad. Sci., III, Sci. Vie 316:404-409(1993).
CC   -!- FUNCTION: Promotes microtubule assembly and stability, and might
CC       be involved in the establishment and maintenance of neuronal
CC       polarity. The C-terminus binds axonal microtubules while the N-
CC       terminus binds neural plasma membrane components, suggesting that
CC       tau functions as a linker protein between both. Axonal polarity is
CC       predetermined by tau localization (in the neuronal cell) in the
CC       domain of the cell body defined by the centrosome. The short
CC       isoforms allow plasticity of the cytoskeleton whereas the longer
CC       isoforms may preferentially play a role in its stabilization.
CC   -!- SUBCELLULAR LOCATION: Mostly found in the axons of neurons, in the
CC       cytosol and in association with plasma membrane components.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=6;
CC         Comment=Additional isoforms seem to exist. Isoforms differ from
CC         each other by the presence or absence of up to 5 of the 14
CC         exons. One of these optional exons contains the additional
CC         tau/MAP repeat. Two different C-termini are obtained either by
CC         the retention or the splicing of intron 13/14;
CC       Name=PNS-Tau;
CC         IsoId=P10637-1; Sequence=Displayed;
CC       Name=Tau-A;
CC         IsoId=P10637-2; Sequence=VSP_003187, VSP_003188;
```

```
CC       Name=Tau-B;
CC         IsoId=P10637-3; Sequence=VSP_003185, VSP_003187, VSP_003188,
CC                         VSP_003189, VSP_003190;
CC       Name=Tau-C;
CC         IsoId=P10637-4; Sequence=VSP_003185, VSP_003187, VSP_003188,
CC                         VSP_003189;
CC       Name=Tau-D;
CC         IsoId=P10637-5; Sequence=VSP_003185, VSP_003187, VSP_003188;
CC       Name=Tau-E;
CC         IsoId=P10637-6; Sequence=VSP_003185, VSP_003186, VSP_003187,
CC                         VSP_003188;
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONS AND AT A LOWER LEVEL IN
CC       THE LIVER AND KIDNEY. PNS-TAU IS EXPRESSED IN THE PERIPHERAL
CC       NERVOUS SYSTEM WHILE THE OTHERS ARE EXPRESSED IN THE CENTRAL
CC       NERVOUS SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: SHORTER FORMS OR LOW MOLECULAR WEIGHT TAU
CC       (LMW-TAU) ARE GENERALLY EXPRESSED AT EARLY DEVELOPMENT STAGES AND
CC       LONGER FORMS OR HIGH MOLECULAR WEIGHT TAU (HMW-TAU) IN THE ADULT
CC       BRAIN.
CC   -!- DOMAIN: The tau/MAP repeat binds to tubulin. Type I isoforms
CC       contain 3 repeats while type II isoforms contain 4 repeats.
CC   -!- PTM: PHOSPHORYLATION AT VARIOUS SERINE AND THREONINE RESIDUES IN
CC       S-P OR T-P MOTIFS BY PROLINE-DIRECTED PROTEIN KINASES (PDPK: CDC2,
CC       CDK5, GSK3, MAPK) (A FEW SITES PER PROTEIN IN INTERPHASE, MORE IN
CC       MITOSIS), AND AT SERINE RESIDUES IN K-X-G-S MOTIFS BY
CC       MAP/MICROTUBULE AFFINITY-REGULATING KINASE (MARK) (BY SIMILARITY).
CC   -!- DISEASE: May be involved in the pathogenesis of cytoplasmic
CC       inclusions (as Mallory bodies) in livers of mice chronically
CC       intoxicated with Griseofulvin or DDC (3,5-diethoxycarbonyl-2,4-
CC       dihydrocollidine), a model for human alcoholic hepatitis.
CC       Alteration of tau (abnormal phosphorylation and crosslinking)
CC       could contribute to Mallory bodies formation and disturbance of
CC       microtubule function in alcoholic liver disease.
CC   -!- SIMILARITY: Contains 4 Tau/MAP repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12914; AAA58343.1; -.
DR   EMBL; U12915; AAA58344.1; -.
DR   EMBL; U12916; AAA58345.1; -.
DR   EMBL; Z12133; CAA78121.1; -.
DR   EMBL; M93266; -; NOT_ANNOTATED_CDS.
DR   EMBL; M18775; AAA40165.1; -.
DR   EMBL; M18776; AAA40166.1; -.
DR   EMBL; D30627; BAA18878.1; -.
DR   PIR; A28820; A28820.
DR   PIR; A45301; A45301.
DR   PIR; B28820; B28820.
DR   MGD; MGI:97180; Mapt.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   InterPro; IPR002955; Tau_protein.
DR   InterPro; IPR001084; Tubulin_Tau.
DR   Pfam; PF00418; tubulin-binding; 4.
DR   PRINTS; PR01261; TAUPROTEIN.
DR   PROSITE; PS00229; TAU_MAP; 4.
KW   Microtubule; Cytoskeleton; Repeat; Alternative splicing; Acetylation;
KW   Phosphorylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   REPEAT      535    565       TAU/MAP MOTIF 1.
FT   REPEAT      566    596       TAU/MAP MOTIF 2.
FT   REPEAT      597    627       TAU/MAP MOTIF 3.
FT   REPEAT      628    659       TAU/MAP MOTIF 4.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   DISULFID    582    613       BY SIMILARITY.
FT   VARSPLIC     33     90       Missing (in isoform Tau-B, isoform Tau-C,
FT                                isoform Tau-D and isoform Tau-E).
FT                                /FTId=VSP_003185.
FT   VARSPLIC     91    112       Missing (in isoform Tau-E).
FT                                /FTId=VSP_003186.
FT   VARSPLIC    113    349       Missing (in isoform Tau-A, isoform Tau-B,
FT                                isoform Tau-C, isoform Tau-D and isoform
FT                                Tau-E).
FT                                /FTId=VSP_003187.
FT   VARSPLIC    367    432       Missing (in isoform Tau-A, isoform Tau-B,
FT                                isoform Tau-C, isoform Tau-D and isoform
FT                                Tau-E).
FT                                /FTId=VSP_003188.
FT   VARSPLIC    566    596       Missing (in isoform Tau-B and isoform
FT                                Tau-C).
FT                                /FTId=VSP_003189.
FT   VARSPLIC    732    732       L -> KAALLSSQVWNYSHDLATITDLGL (in isoform
FT                                Tau-B).
FT                                /FTId=VSP_003190.
FT   CONFLICT      2      2       D -> N (IN REF. 1).
FT   CONFLICT      8      8       D -> N (IN REF. 1).
FT   CONFLICT    527    527       P -> T (IN REF. 2; CAA78121).
FT   CONFLICT    671    671       E -> Q (IN REF. 1).
SQ   SEQUENCE   732 AA;  76112 MW;  BFDF0767E41C7A3A CRC64;

  Query Match             76.8%;  Score 43;  DB 1;  Length 732;
  Best Local Similarity   77.8%;  Pred. No. 63;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | | |||||
Db        468 PSPKTPPGS 476


RESULT 4
TAU_RAT
ID   TAU_RAT        STANDARD;      PRT;   751 AA.
AC   P19332; Q63567; Q63677; Q9QW06;
DT   01-NOV-1990 (Rel. 16, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Microtubule-associated protein tau (Neurofibrillary tangle protein)
DE   (Paired helical filament-tau) (PHF-tau).
GN   MAPT OR MTAPT OR TAU.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM TAU-B).
RC   TISSUE=Pheochromocytoma;
RX   MEDLINE=92179305; PubMed=1542696;
RA   Goedert M., Spillantini M.G., Crowther R.A.;
RT   "Cloning of a big tau microtubule-associated protein characteristic of
RT   the peripheral nervous system.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:1983-1987(1992).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM TAU-B).
RC   TISSUE=Dorsal root ganglion;
RX   MEDLINE=94013081; PubMed=8408300;
RA   Georgieff I.S., Liem R.K.H., Couchie D., Mavilia C., Nunez J.,
RA   Shelanski M.L.;
RT   "Expression of high molecular weight tau in the central and peripheral
RT   nervous systems.";
RL   J. Cell Sci. 105:729-737(1993).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM TAU-F).
RC   STRAIN=Wistar; TISSUE=Brain;
RX   MEDLINE=94334997; PubMed=8057376;
RA   Sadot E., Marx R., Barg J., Behar L., Ginzburg I.;
RT   "Complete sequence of 3'-untranslated region of tau from rat central
RT   nervous system. Implications for mRNA heterogeneity.";
RL   J. Mol. Biol. 241:325-331(1994).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS TAU-E AND TAU-G).
RC   TISSUE=Brain;
```

```
RX   MEDLINE=90180457; PubMed=2560640;
RA   Kosik K.S., Orecchio L.D., Bakalis S., Neve R.L.;
RT   "Developmentally regulated expression of specific tau sequences.";
RL   Neuron 2:1389-1397(1989).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORMS TAU-E AND TAU-C).
RX   MEDLINE=89359509; PubMed=2504728;
RA   Kanai Y., Takemura R., Oshima T., Mori H., Ihara Y., Yanagisawa M.,
RA   Masaki T., Hirokawa N.;
RT   "Expression of multiple tau isoforms and microtubule bundle formation
RT   in fibroblasts transfected with a single tau cDNA.";
RL   J. Cell Biol. 109:1173-1184(1989).
RN   [6]
RP   SEQUENCE OF 359-460 FROM N.A. (ISOFORM TAU-A), AND SEQUENCE OF 105-112
RP   AND 367-460 FROM N.A. (ISOFORM TAU-D).
RC   TISSUE=Spinal cord;
RX   MEDLINE=95054048; PubMed=7964751;
RA   Mavilia C., Couchie D., Nunez J.;
RT   "Diversity of high-molecular-weight tau proteins in different regions
RT   of the nervous system.";
RL   J. Neurochem. 63:2300-2306(1994).
RN   [7]
RP   SEQUENCE OF 696-751 FROM N.A. (ISOFORMS TAU-A; TAU-B; TAU-C; TAU-D;
RP   TAU-E; TAU-F AND TAU-G), AND SEQUENCE OF 751-774 FROM N.A. (ISOFORM
RP   TAU-H).
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=95182802; PubMed=7877441;
RA   Sawa A., Oyama F., Matsushita M., Ihara Y.;
RT   "Molecular diversity at the carboxyl terminus of human and rat tau.";
RL   Brain Res. Mol. Brain Res. 27:111-117(1994).
CC   -!- FUNCTION: Promotes microtubule assembly and stability, and might
CC       be involved in the establishment and maintenance of neuronal
CC       polarity. The C-terminus binds axonal microtubules while the N-
CC       terminus binds neural plasma membrane components, suggesting that
CC       tau functions as a linker protein between both. Axonal polarity is
CC       predetermined by tau localization (in the neuronal cell) in the
CC       domain of the cell body defined by the centrosome. The short
CC       isoforms allow plasticity of the cytoskeleton whereas the longer
CC       isoforms may preferentially play a role in its stabilization.
CC   -!- SUBCELLULAR LOCATION: Mostly found in the axons of neurons, in the
CC       cytosol and in association with plasma membrane components.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=8;
CC         Comment=Additional isoforms seem to exist. Isoforms differ from
CC         each other by the presence or absence of up to 4 of the 14
CC         exons. Two different C-termini are obtained either by the
CC         retention or the splicing of intron 13/14;
CC       Name=Tau-A; Synonyms=SC1;
CC         IsoId=P19332-1; Sequence=Displayed;
CC       Name=Tau-B; Synonyms=Big-tau, HMW-tau;
CC         IsoId=P19332-2; Sequence=VSP_003194;
CC       Name=Tau-C;
CC         IsoId=P19332-3; Sequence=VSP_003192, VSP_003193, VSP_003194;
CC       Name=Tau-D; Synonyms=SC2;
CC         IsoId=P19332-4; Sequence=VSP_003193;
CC       Name=Tau-E;
CC         IsoId=P19332-5; Sequence=VSP_003193, VSP_003194;
CC       Name=Tau-F;
CC         IsoId=P19332-6; Sequence=VSP_003191, VSP_003193, VSP_003194;
CC       Name=Tau-G; Synonyms=Fetal-tau;
CC         IsoId=P19332-7; Sequence=VSP_003192, VSP_003193, VSP_003194,
CC                                  VSP_003195;
CC       Name=Tau-H;
CC         IsoId=P19332-8; Sequence=VSP_003196;
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONS. THE LARGER FORMS (TAU-A
CC       AND TAU-B) ARE PREFERENTIALLY EXPRESSED IN THE PERIPHERAL NERVOUS
CC       SYSTEM WHILE THE OTHERS ARE EXPRESSED IN THE CENTRAL NERVOUS
CC       SYSTEM. LOW AMOUNTS OF THE LARGER FORMS ARE ALSO FOUND IN LIMITED
CC       AREAS OF THE CNS.
CC   -!- DEVELOPMENTAL STAGE: DURING THE IMMEDIATE POSTNATAL PERIOD, THE
CC       DORSAL ROOT GANGLIA EXPRESS ALL ISOFORMS WHEREAS ONLY THE LARGER
CC       FORMS PERSIST IN THE ADULTS.
CC   -!- DOMAIN: The tau/MAP repeat binds to tubulin. Type I isoforms
CC       contain 3 repeats while type II isoforms contain 4 repeats.
CC   -!- PTM: PHOSPHORYLATION AT VARIOUS SERINE AND THREONINE RESIDUES IN
CC       S-P OR T-P MOTIFS BY PROLINE-DIRECTED PROTEIN KINASES (PDPK: CDC2,
CC       CDK5, GSK3, MAPK) (A FEW SITES PER PROTEIN IN INTERPHASE, MORE IN
CC       MITOSIS), AND AT SERINE RESIDUES IN K-X-G-S MOTIFS BY
CC       MAP/MICROTUBULE AFFINITY-REGULATING KINASE (MARK) (BY SIMILARITY).
CC   -!- SIMILARITY: Contains 4 Tau/MAP repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M84156; AAA42204.1; -.
DR   EMBL; X79321; CAA55889.1; -.
DR   EMBL; D30628; -; NOT_ANNOTATED_CDS.
DR   EMBL; D30629; -; NOT_ANNOTATED_CDS.
DR   PIR; A38235; A38235.
DR   PIR; JS0306; JS0306.
DR   InterPro; IPR002955; Tau_protein.
DR   InterPro; IPR001084; Tubulin_Tau.
DR   Pfam; PF00418; tubulin-binding; 4.
DR   PRINTS; PR01261; TAUPROTEIN.
DR   PROSITE; PS00229; TAU_MAP; 4.
KW   Microtubule; Cytoskeleton; Repeat; Alternative splicing; Acetylation;
KW   Phosphorylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   REPEAT      554    584       TAU/MAP MOTIF 1.
FT   REPEAT      585    615       TAU/MAP MOTIF 2.
FT   REPEAT      616    646       TAU/MAP MOTIF 3.
FT   REPEAT      647    678       TAU/MAP MOTIF 4.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   DISULFID    601    632       BY SIMILARITY.
FT   VARSPLIC     33     90       Missing (in isoform Tau-F).
FT                                /FTId=VSP_003191.
FT   VARSPLIC     62     90       Missing (in isoform Tau-C and isoform
FT                                Tau-G).
FT                                /FTId=VSP_003192.
FT   VARSPLIC    113    366       Missing (in isoform Tau-C, isoform Tau-D,
FT                                isoform Tau-E, isoform Tau-F and isoform
FT                                Tau-G).
FT                                /FTId=VSP_003193.
FT   VARSPLIC    386    451       Missing (in isoform Tau-B, isoform Tau-C,
FT                                isoform Tau-E, isoform Tau-F and isoform
FT                                Tau-G).
FT                                /FTId=VSP_003194.
FT   VARSPLIC    585    615       Missing (in isoform Tau-G).
FT                                /FTId=VSP_003195.
FT   VARSPLIC    751    751       L -> KPVLLSSEVWNYSHDFGHHTDLGL (in isoform
FT                                Tau-H).
FT                                /FTId=VSP_003196.
FT   CONFLICT    254    254       F -> L (IN REF. 2).
FT   CONFLICT    283    283       G -> A (IN REF. 2).
FT   CONFLICT    291    291       H -> D (IN REF. 2).
FT   CONFLICT    617    617       H -> Q (IN REF. 2, 3 AND 4).
FT   CONFLICT    704    704       Y -> H (IN REF. 3).
FT   CONFLICT    733    733       P -> A (IN REF. 2).
SQ   SEQUENCE   751 AA;  78432 MW;  B96B7329444D44B2 CRC64;

  Query Match             76.8%;  Score 43;  DB 1;  Length 751;
  Best Local Similarity   77.8%;  Pred. No. 65;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | | |||||
Db        487 PSPKTPPGS 495


RESULT 5
YGHS_ECOLI
```

```
ID   YGHS_ECOLI     STANDARD;      PRT;   237 AA.
AC   Q46843;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical ATP-binding protein yghS.
GN   YGHS OR B2985.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SIMILARITY: TO E.COLI YGHR AND YGHT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U28377; AAA69152.1; -.
DR   EMBL; AE000381; AAC76021.1; -.
DR   PIR; G65084; G65084.
DR   EcoGene; EG13003; yghS.
KW   Hypothetical protein; ATP-binding; Complete proteome.
FT   NP_BIND      21     28       ATP (POTENTIAL).
SQ   SEQUENCE   237 AA;  26346 MW;  69D8AE6673D7DA6F CRC64;

  Query Match             75.0%;  Score 42;  DB 1;  Length 237;
  Best Local Similarity   87.5%;  Pred. No. 28;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KPSTPPGS 9
              |||||||:
Db         84 KPSTPPGN 91


RESULT 6
GUN2_THEFU
ID   GUN2_THEFU     STANDARD;      PRT;   441 AA.
AC   P26222;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Endoglucanase E-2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-2)
DE   (Cellulase E-2) (Cellulase E2).
GN   CELB.
OS   Thermomonospora fusca.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX   NCBI_TaxID=2021;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YX;
RX   MEDLINE=91258320; PubMed=1904434;
RA   Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT   "DNA sequences of three beta-1,4-endoglucanase genes from
RT   Thermomonospora fusca.";
RL   J. Bacteriol. 173:3397-3407(1991).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RC   STRAIN=YX;
RA   Jung E.D., Lao G., Irwin D., Barr B., Benjamin A., Wilson D.B.;
RL   Submitted (MAY-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 32-47.
RA   Wilson D.B.;
RT   "Cellulases of Thermomonospora fusca.";
RL   Meth. Enzymol. 160:314-323(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 32-317.
RX   MEDLINE=94002001; PubMed=8399160;
RA   Spezio M., Wilson D.B., Karplus P.A.;
RT   "Crystal structure of the catalytic domain of a thermophilic
RT   endocellulase.";
RL   Biochemistry 32:9906-9916(1993).
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose, lichenin and cereal beta-D-glucans.
CC   -!- PATHWAY: Cellulose degradation.
CC   -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC       domain.
CC   -!- SIMILARITY: Belongs to cellulase family B (family 6 of glycosyl
CC       hydrolases).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73321; AAC06388.1; -.
DR   PIR; A42360; A42360.
DR   PIR; T12011; T12011.
DR   PDB; 1TML; 31-JAN-94.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR001524; Glyco_hydro_6.
DR   Pfam; PF00553; CBM_2; 1.
DR   Pfam; PF01341; Glyco_hydro_6; 1.
DR   PRINTS; PR00733; GLHYDRLASE6.
DR   ProDom; PD003733; Glyco_hydro_6; 1.
DR   SMART; SM00637; CBD_II; 1.
DR   PROSITE; PS00561; CBD_BACTERIAL; 1.
DR   PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR   PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT   SIGNAL       1     31
FT   CHAIN       32    441       ENDOGLUCANASE E-2.
FT   DOMAIN      32    320       CATALYTIC.
FT   DOMAIN     321    340       LINKER.
FT   DOMAIN     341    441       CELLULOSE-BINDING.
FT   ACT_SITE   110    110
FT   ACT_SITE   148    148       PROTON DONOR.
FT   ACT_SITE   296    296       NUCLEOPHILE.
FT   DISULFID   111    156
FT   DISULFID   263    298
FT   DISULFID   346    438       POTENTIAL.
FT   STRAND      36     36
FT   TURN        40     41
FT   HELIX       43     50
FT   TURN        52     53
FT   TURN        55     56
FT   HELIX       57     63
FT   TURN        64     66
FT   STRAND      69     69
FT   STRAND      71     73
FT   HELIX       78     95
FT   TURN        96     96
FT   STRAND      98     98
FT   STRAND     100    103
FT   TURN       109    112
FT   HELIX      122    134
FT   TURN       135    138
```

```
FT   STRAND      142    145
FT   TURN        147    148
FT   HELIX       149    153
FT   TURN        154    155
FT   HELIX       158    178
FT   TURN        180    181
FT   STRAND      183    187
FT   HELIX       196    205
FT   TURN        206    207
FT   HELIX       208    211
FT   STRAND      214    217
FT   TURN        219    220
FT   HELIX       225    239
FT   TURN        240    240
FT   TURN        242    243
FT   STRAND      245    249
FT   TURN        254    255
FT   TURN        259    260
FT   TURN        266    267
FT   STRAND      275    275
FT   TURN        281    282
FT   STRAND      283    288
FT   TURN        292    293
FT   STRAND      294    294
FT   TURN        302    303
FT   STRAND      305    305
FT   HELIX       307    315
FT   TURN        316    316
SQ   SEQUENCE   441 AA;  45843 MW;  87218E4537092AE5 CRC64;

  Query Match             73.2%;  Score 41;  DB 1;  Length 441;
  Best Local Similarity   70.0%;  Pred. No. 72;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              | |: |||||
Db        334 PTPTPPPGSS 343


RESULT 7
SYT5_HUMAN
ID   SYT5_HUMAN     STANDARD;      PRT;   386 AA.
AC   O00445; Q86X72;
DT   01-NOV-1997 (Rel. 35, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Synaptotagmin V (SytV).
GN   SYT5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97321058; PubMed=9177789;
RA   Craxton M.A., Olsen A., Goedert M.;
RT   "Human synaptotagmin V (SYT5): sequence, genomic structure, and
RT   chromosomal location.";
RL   Genomics 42:165-169(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: May be involved in Ca(2+)-dependent exocytosis of
CC       secretory vesicles through Ca(2+) and phospholipid binding to the
CC       C2 domain or may serve as Ca(2+) sensors in the process of
CC       vesicular trafficking and exocytosis (By similarity).
CC   -!- SUBUNIT: Homodimer. Can also form heterodimers.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Synaptic
CC       vesicles.
CC   -!- SIMILARITY: Belongs to the synaptotagmin family.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X96783; CAA65579.1; -.
DR   EMBL; BC046157; AAH46157.1; -.
DR   HSSP; P21707; 1BYN.
DR   Genew; HGNC:11513; SYT5.
DR   MIM; 600782; -.
DR   GO; GO:0007268; P:synaptic transmission; TAS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR002149; LRI.
DR   InterPro; IPR001565; Synaptotagmin.
DR   Pfam; PF00168; C2; 2.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 2.
DR   PROSITE; PS00499; C2_DOMAIN_1; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
KW   Transmembrane; Repeat; Synapse.
FT   DOMAIN        1     24       VESICULAR (POTENTIAL).
FT   TRANSMEM     25     45       POTENTIAL.
FT   DOMAIN       46    386       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      124    211       C2 DOMAIN 1.
FT   DOMAIN      253    344       C2 DOMAIN 2.
FT   CONFLICT    111    111       R -> Q (IN REF. 2).
SQ   SEQUENCE   386 AA;  42900 MW;  96A36792D177FD55 CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 386;
  Best Local Similarity   70.0%;  Pred. No. 86;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              | | ||| ||
Db         11 PSPDTPPDSS 20


RESULT 8
TRZA_RHOCO
ID   TRZA_RHOCO     STANDARD;      PRT;   476 AA.
AC   Q52725;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   S-triazine hydrolase (EC 3.8.1.-) (N-ethylammeline chlorohydrolase).
GN   TRZA.
OS   Rhodococcus corallinus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
```

```
OC   Corynebacterineae; Gordoniaceae; Gordonia.
OX   NCBI_TaxID=36822;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=NRRL 15444R;
RX   MEDLINE=96011356; PubMed=7592318;
RA   Shao Z.Q., Seffens W., Mulbry W., Behki R.M.;
RT   "Cloning and expression of the s-triazine hydrolase gene (trzA) from
RT   Rhodococcus corallinus and development of Rhodococcus recombinant
RT   strains capable of dealkylating and dechlorinating the herbicide
RT   atrazine.";
RL   J. Bacteriol. 177:5748-5755(1995).
CC   -!- FUNCTION: HYDROLYTIC DEAMINATION OF THE S-TRIAZINE SUBSTRATE
CC       MELAMINE.
CC   -!- PATHWAY: Melamine degradation pathway; first step.
CC   -!- SIMILARITY: Belongs to the ATZ/TRZ family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L16534; AAA90931.1; -.
DR   PIR; T46666; T46666.
DR   InterPro; IPR006680; Amidohydro_1.
DR   Pfam; PF01979; Amidohydro_1; 1.
KW   Hydrolase.
FT   INIT_MET      0      0
SQ   SEQUENCE   476 AA;  50727 MW;  64D953DB2E92C73E CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 476;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPG 8
              || |||||
Db         46 PKSSTPPG 53


RESULT 9
APG_ARATH
ID   APG_ARATH      STANDARD;      PRT;   534 AA.
AC   P40602; Q93Z14; Q9LNT8;
DT   01-FEB-1995 (Rel. 31, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Anter-specific proline-rich protein APG precursor.
GN   APG OR AT1G20130 OR T20H2.9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94004980; PubMed=8401599;
RA   Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,
RA   Draper J., Scott R.;
RT   "Gametophytic and sporophytic expression of an anther-specific
RT   Arabidopsis thaliana gene.";
RL   Plant J. 3:111-120(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=22954850; PubMed=14593172;
RA   Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA   Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA   Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA   Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA   Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA   Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA   Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA   Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA   Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA   Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA   Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Empirical analysis of transcriptional activity in the Arabidopsis
RT   genome.";
RL   Science 302:842-846(2003).
CC   -!- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GAMETOPHYTIC CELL
CC       TYPES IN THE ANTHER, ONLY IN MALE FERTILE PLANTS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
CC       MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
CC       MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC       MATURATION.
CC   -!- SIMILARITY: Belongs to the "GDSL" family of lipolytic enzymes.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC       gene model prediction.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X60377; CAA42925.1; -.
DR   EMBL; AC022472; AAF79900.1; ALT_SEQ.
DR   EMBL; AY058847; AAL24235.1; -.
DR   PIR; S21961; S21961.
DR   InterPro; IPR001087; Lipase_GDSL.
DR   InterPro; IPR008265; Lipase_GDSL_AS.
DR   Pfam; PF00657; Lipase_GDSL; 1.
DR   PROSITE; PS01098; LIPASE_GDSL_SER; 1.
KW   Signal.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36    534       ANTER-SPECIFIC PROLINE-RICH PROTEIN APG.
FT   ACT_SITE    211    211       BY SIMILARITY.
FT   ACT_SITE    511    511       POTENTIAL.
FT   CONFLICT     77     77       S -> P (IN REF. 1).
FT   CONFLICT    141    141       E -> A (IN REF. 1).
FT   CONFLICT    325    325       E -> H (IN REF. 1).
SQ   SEQUENCE   534 AA;  58007 MW;  BA851DC3CF7429DB CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 534;
  Best Local Similarity   70.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy          1 PKPSTPPGSS 10
              |||  ||| |
Db         67 PKPVAPPGPS 76


RESULT 10
COE2_BRARE
ID   COE2_BRARE     STANDARD;      PRT;   579 AA.
AC   O93375;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transcription factor COE2.
GN   COE2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99002949; PubMed=9784615;
RA   Bally-Cuif L., Dubois L., Vincent A.;
RT   "Molecular cloning of Zcoe2, the zebrafish homolog of Xenopus Xcoe2
RT   and mouse EBF-2, and its expression during primary neurogenesis.";
RL   Mech. Dev. 77:85-90(1998).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: First detected at the 75% epiboly stage where
CC       it covers the anterior neural plate. Widely expressed in the
CC       presumptive mesencephalon and rhombomeres 1-4 until the 2-3-somite
CC       stage, with expression persisting in ngn1-positive clusters. First
CC       detected in the olfactory placodes at the 5-somite stage. In the
CC       spinal cord, detected in ngn1-positive clusters of primary
CC       neuroblasts during the early somite stages. Expression decreases
CC       in the spinal cord from the 30-somite stage but persists in the
CC       olfactory bulb and regions of the rhombencephalon and brain.
CC   -!- SIMILARITY: Belongs to the COE family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072657; AAC96103.1; -.
DR   ZFIN; ZDB-GENE-990715-11; coe2.
DR   InterPro; IPR001092; HLH_basic.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR003523; TF_COE.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00353; HLH; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01345; COE; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   Developmental protein; Zinc-finger.
FT   ZN_FING     149    168       C5-TYPE (POTENTIAL).
FT   DOMAIN      460    537       SER/THR/PRO-RICH.
SQ   SEQUENCE   579 AA;  63528 MW;  E2DD1919055EEA43 CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 579;
  Best Local Similarity   87.5%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 PSTPPGSS 10
              ||:|||||
Db        517 PSSPPGSS 524


RESULT 11
FZD9_MOUSE
ID   FZD9_MOUSE     STANDARD;      PRT;   592 AA.
AC   Q9R216; O35494; Q9CX16; Q9R2B3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Frizzled 9 precursor (Frizzled-9) (Fz-9) (mFz9) (mFz3).
GN   FZD9 OR FZD3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
RC   STRAIN=129/SvJ; TISSUE=Brain;
RX   MEDLINE=99216417; PubMed=10198163;
RA   Wang Y.-K., Spoerle R., Paperna T., Schughart K., Francke U.;
RT   "Characterization and expression pattern of the frizzled Gene Fzd9,
RT   the mouse homolog of FZD9 which is deleted in Williams-Beuren
RT   syndrome.";
RL   Genomics 57:235-248(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE OF 44-592 FROM N.A.
RA   Van Raay T.J., Rasmussen J.T., Rao M.S.;
RT   "A novel mouse frizzled gene expressed in early neural development.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 144-592 FROM N.A.
RC   STRAIN=BALB/c;
RA   Calo L., Mimmack M.L., Keverne E.B., Emson P.C.;
RT   "Localization of the mouse frizzled gene mFZD3 in the olfactory
RT   epithelium and in the vomeronasal organ.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Receptor for Wnt proteins. Most of frizzled receptors
CC       are coupled to the beta-catenin canonical signaling pathway, which
CC       leads to the activation of dishevelled proteins, inhibition of
CC       GSK-3 kinase, nuclear accumulation of beta-catenin and activation
CC       of Wnt target genes. A second signaling pathway involving PKC and
CC       calcium fluxes has been seen for some family members, but it is
CC       not yet clear if it represents a distinct pathway or if it can be
CC       integrated in the canonical pathway, as PKC seems to be required
CC       for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem
CC       to involve interactions with G-proteins. May be involved in
CC       transduction and intercellular transmission of polarity
CC       information during tissue morphogenesis and/or in differentiated
CC       tissues.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Localizes in the
CC       plasma membrane.
CC   -!- TISSUE SPECIFICITY: In the embryo, found in the neural tube, trunk
CC       skeletal muscle precursors (myotomes), limb skeletal anlagen,
```

```
CC       craniofacial regions and nephric ducts. In the adult, expression
CC       is abundant in heart, brain, testis and skeletal muscle. In the
CC       testis, expressed in all spermatogenic cell types. Lower levels in
CC       adult lung, liver and kidney. Barely detectable in spleen.
CC       Expressed also in chondrocytes.
CC   -!- DEVELOPMENTAL STAGE: Not detected at embryonic day 7 (E7), weakly
CC       at E11 and strongly at E15 and E17. Expression covers the entire
CC       neural tube at 9.5 dpc, decreases at 10.5 dpc and becomes
CC       detectable only in the lumbar to tail regions at 11.5 dpc. In the
CC       somites, expression begins at 10.5 dpc to become upregulated all
CC       along the rostrocaudal trunk axis at 11.5 dpc. In craniofacial
CC       teritories, expression is first detected at 11.5 dpc in restricted
CC       areas of the nose, the maxillar mandibular and second branchial
CC       arch anlagen. At 11.5 dpc, predominantly expressed in restricted
CC       areas of the nose, dorsally to the eye and in the caudal
CC       pharyngeal region.
CC   -!- DOMAIN: Lys-Thr-X-X-X-Trp motif is involved in the activation of
CC       the Wnt/beta-catenin signaling pathway (By similarity).
CC   -!- DOMAIN: The FZ domain is involved in binding with Wnt ligands (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Fz/Smo G-protein coupled receptor
CC       family.
CC   -!- SIMILARITY: Contains 1 frizzled (FZ) domain.
CC   -!- CAUTION: Has been first described as FZD3 in litterature.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF088850; AAD27789.1; -.
DR   EMBL; AK021164; BAB32311.1; -.
DR   EMBL; AF033585; AAB87503.2; -.
DR   EMBL; Y17709; CAB44237.1; -.
DR   MGD; MGI:1313278; Fzd9.
DR   InterPro; IPR000539; Frizzled.
DR   InterPro; IPR000024; Fz_domain.
DR   InterPro; IPR000832; GPCR_secretin.
DR   Pfam; PF01534; Frizzled; 1.
DR   Pfam; PF01392; Fz; 1.
DR   PRINTS; PR00489; FRIZZLED.
DR   SMART; SM00063; FRI; 1.
DR   PROSITE; PS50038; FZ; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   Multigene family; G-protein coupled receptor; Transmembrane;
KW   Developmental protein; Wnt signaling pathway; Glycoprotein; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    592       FRIZZLED 9.
FT   DOMAIN       24    230       EXTRACELLULAR.
FT   TRANSMEM    231    251       1 (POTENTIAL).
FT   DOMAIN      252    267       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    268    288       2 (POTENTIAL).
FT   DOMAIN      289    316       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    317    337       3 (POTENTIAL).
FT   DOMAIN      338    356       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    357    377       4 (POTENTIAL).
FT   DOMAIN      378    401       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    402    422       5 (POTENTIAL).
FT   DOMAIN      423    448       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    449    469       6 (POTENTIAL).
FT   DOMAIN      470    509       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    510    530       7 (POTENTIAL).
FT   DOMAIN      531    592       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       35    156       FZ.
FT   SITE        533    538       LYS-THR-X-X-X-TRP MOTIF.
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     66     66       S -> P (IN REF. 3).
FT   CONFLICT     73     74       QL -> HC (IN REF. 2).
FT   CONFLICT     93     93       L -> F (IN REF. 2).
FT   CONFLICT    144    144       P -> S (IN REF. 4).
FT   CONFLICT    221    221       E -> K (IN REF. 4).
FT   CONFLICT    237    237       A -> P (IN REF. 4).
FT   CONFLICT    308    308       G -> D (IN REF. 3).
FT   CONFLICT    374    374       V -> F (IN REF. 4).
FT   CONFLICT    592    592       L -> P (IN REF. 2).
SQ   SEQUENCE   592 AA;  64994 MW;  21B2D4F8CE232965 CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 592;
  Best Local Similarity   60.0%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              |:|: ||| |
Db        178 PRPARPPGDS 187


RESULT 12
3BP1_MOUSE
ID   3BP1_MOUSE     STANDARD;      PRT;   601 AA.
AC   P55194; Q99KK8;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   SH3-domain binding protein 1 (3BP-1).
GN   SH3BP1 OR 3BP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95347339; PubMed=7621827;
RA   Cicchetti P., Ridley A.J., Zheng Y., Cerione R.A., Baltimore D.;
RT   "3BP-1, an SH3 domain binding protein, has GAP activity for Rac and
RT   inhibits growth factor-induced membrane ruffling in fibroblasts.";
RL   EMBO J. 14:3127-3135(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 263-601 FROM N.A.
RX   MEDLINE=92358242; PubMed=1379745;
RA   Cicchetti P., Mayer B.J., Thiel G., Baltimore D.;
RT   "Identification of a protein that binds to the SH3 region of Abl and
RT   is similar to Bcr and GAP-rho.";
RL   Science 257:803-806(1992).
CC   -!- FUNCTION: Binds differentially to the SH3 domains of certain
CC       proteins of signal transduction pathways. This protein binds
CC       preferentially to c-Abl proto-oncogene, SRC and GRB2. Shows GAP
CC       activity for Rac-related proteins but not for Rho- or Ras-related
CC       proteins. It inhibits PDGF-induced membrane ruffling mediated by
CC       Rac.
```

```
CC   -!- TISSUE SPECIFICITY: Expressed in all tissues examined. Highest
CC       levels found in spleen and brain, lowest in heart and liver.
CC   -!- SIMILARITY: Contains 1 Rho-GAP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X87671; CAA61011.1; -.
DR   EMBL; BC004598; AAH04598.1; -.
DR   PIR; S56144; S56144.
DR   PDB; 1ABO; 15-OCT-95.
DR   MGD; MGI:104603; Sh3bp1.
DR   InterPro; IPR008936; Rho_GAP.
DR   InterPro; IPR000198; RhoGAP.
DR   Pfam; PF00620; RhoGAP; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
KW   GTPase activation; SH3-binding; 3D-structure.
FT   DOMAIN      196    390       RHO-GAP.
FT   SITE        529    537       SH3-BINDING.
FT   DOMAIN      113    117       POLY-GLU.
FT   CONFLICT     46     48       RSG -> PLS (IN REF. 2).
FT   CONFLICT    229    229       MISSING (IN REF. 2).
FT   CONFLICT    261    261       G -> D (IN REF. 2).
FT   CONFLICT    592    593       PA -> RP (IN REF. 2).
SQ   SEQUENCE   601 AA;  65260 MW;  0FBBF357EEB02ECE CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 601;
  Best Local Similarity   77.8%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | || ||||
Db        500 PAPSLPPGS 508


RESULT 13
ACK1_HUMAN
ID   ACK1_HUMAN     STANDARD;      PRT;  1036 AA.
AC   Q07912; Q8N6U7; Q96H59;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Activated CDC42 kinase 1 (EC 2.7.1.112) (ACK-1).
GN   ACK1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH CDC42.
RC   TISSUE=Hippocampus;
RX   MEDLINE=93268389; PubMed=8497321;
RA   Manser E., Leung T., Salihuddin H., Tan L., Lim L.;
RT   "A non-receptor tyrosine kinase that inhibits the GTPase activity of
RT   p21cdc42.";
RL   Nature 363:364-367(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain, and Uterus;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   STRUCTURE BY NMR OF 448-489.
RX   MEDLINE=99287324; PubMed=10360579;
RA   Mott H.R., Owen D., Nietlispach D., Lowe P.N., Manser E., Lim L.,
RA   Laue E.D.;
RT   "Structure of the small G protein Cdc42 bound to the GTPase-binding
RT   domain of ACK.";
RL   Nature 399:384-388(1999).
CC   -!- FUNCTION: Tyrosine kinase, that after binding to CDC42, inhibits
CC       both its intrinsic and stimulated GTPase activity.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: Interacts with CDC42.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q07912-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q07912-2; Sequence=VSP_008655, VSP_008656;
CC         Note=No experimental confirmation available;
CC   -!- SIMILARITY: Belongs to the Tyr family of protein kinases.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 1 CRIB domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13738; AAA53570.2; -.
DR   EMBL; BC008884; AAH08884.1; -.
DR   EMBL; BC028164; AAH28164.1; -.
DR   PIR; S33596; S33596.
DR   PDB; 1CF4; 27-JUN-01.
DR   MIM; 606994; -.
DR   GO; GO:0005095; F:GTPase inhibitor activity; TAS.
DR   GO; GO:0004715; F:non-membrane spanning protein tyrosine kina. . .; TAS.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; TAS.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   InterPro; IPR000449; UBA_domain.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00627; UBA; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50002; SH3; 1.
DR   PROSITE; PS50108; CRIB; FALSE_NEG.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; SH3 domain;
KW   3D-structure; Alternative splicing.
```

```
FT   DOMAIN       126    385       PROTEIN KINASE.
FT   DOMAIN       386    448       SH3.
FT   DOMAIN       454    466       CRIB.
FT   DOMAIN       577    956       PRO-RICH.
FT   NP_BIND      132    140       ATP (BY SIMILARITY).
FT   BINDING      158    158       ATP (BY SIMILARITY).
FT   ACT_SITE     252    252       BY SIMILARITY.
FT   VARSPLIC     485    528       LYLGNPMDPPDLLSVELSTSRPPQHLGGVKKPTYDPVSEDQ
FT                                 DPL -> CPFSAFSPGHPPAETCGQVLWTGRREACASDPRL
FT                                 HPVSSRTKGL (in isoform 2).
FT                                 /FTId=VSP_008655.
FT   VARSPLIC     529   1036       Missing (in isoform 2).
FT                                 /FTId=VSP_008656.
FT   CONFLICT     138    138       G -> V (IN REF. 2; AAH08884).
FT   CONFLICT     304    352       TRTFSHASDTWMFGVTLWEMFTYGQEPWIGLNGSQILHKID
FT                                 KEGERLPR -> PPWRDISASSSTQFPHAVPCFPTSLLAKL
FT                                 LLRHSVPASSREIKLVSILC (IN REF. 2;
FT                                 AAH08884).
FT   CONFLICT     353   1036       Missing (IN REF. 2; AAH08884).
SQ   SEQUENCE   1036 AA;  114327 MW;  B9B90BA7E3E22DFF CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 1036;
  Best Local Similarity   70.0%;  Pred. No. 2.4e+02;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              | |  |||||
Db        791 PSPLVPPGSS 800


RESULT 14
ACK1_MOUSE
ID   ACK1_MOUSE     STANDARD;      PRT;  1055 AA.
AC   O54967; Q8C2U0; Q8K0K4;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Activated CDC42 kinase 1 (EC 2.7.1.112) (ACK-1) (Non-receptor protein
DE   tyrosine kinase Ack) (Tyrosine kinase non-receptor protein 2).
GN   ACK1 OR TNK2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RA   Her J.-H., Bolen J.B.;
RT   "The protein tyrosine kinase Ack is associated with and activated in
RT   vivo by CDC42Hs.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RC   STRAIN=C57BL/6; TISSUE=Brain, and Colon;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 121-1055 FROM N.A. (ISOFORM 3).
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: Tyrosine kinase, that after binding to CDC42, inhibits
CC       both its intrinsic and stimulated GTPase activity (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: Interacts with CDC42 (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=O54967-1; Sequence=Displayed;
CC         Note=No experimental confirmation available;
CC       Name=2;
CC         IsoId=O54967-2; Sequence=VSP_008657, VSP_008658;
CC         Note=No experimental confirmation available;
CC       Name=3;
CC         IsoId=O54967-3; Sequence=VSP_008657;
CC   -!- SIMILARITY: Belongs to the Tyr family of protein kinases.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 1 CRIB domain.
CC   -!- CAUTION: Ref.2 (AAH31168) sequence differs from that shown due to
CC       the presence of a sequence of unknown origin.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF037260; AAC04786.1; -.
DR   EMBL; BC031168; AAH31168.1; ALT_SEQ.
DR   EMBL; BC052421; AAH52421.1; -.
DR   EMBL; AK087965; BAC40063.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:1858308; Tnk2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001245; Tyr_pkinase.
```

```
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   InterPro; IPR000449; UBA_domain.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00627; UBA; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; SH3 domain;
KW   Alternative splicing.
FT   DOMAIN      126    385       PROTEIN KINASE.
FT   DOMAIN      386    448       SH3.
FT   DOMAIN      454    466       CRIB.
FT   DOMAIN      517    950       PRO-RICH.
FT   NP_BIND     132    140       ATP (BY SIMILARITY).
FT   BINDING     158    158       ATP (BY SIMILARITY).
FT   ACT_SITE    252    252       BY SIMILARITY.
FT   VARSPLIC    515    531       REPPPRPPQPAIFTQKT -> KP (in isoform 2 and
FT                                isoform 3).
FT                                /FTId=VSP_008657.
FT   VARSPLIC    980   1011       Missing (in isoform 2).
FT                                /FTId=VSP_008658.
FT   CONFLICT     57     58       SG -> RR (IN REF. 2).
FT   CONFLICT    649    649       V -> A (IN REF. 2 AND 3).
FT   CONFLICT    818    818       L -> V (IN REF. 3).
FT   CONFLICT    955    955       A -> T (IN REF. 2; AAH52421 AND 3).
SQ   SEQUENCE   1055 AA;  116835 MW;  FBC9DACD85B2003F CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 1055;
  Best Local Similarity   70.0%;  Pred. No. 2.4e+02;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              | |  |||||
Db        807 PSPLVPPGSS 816


RESULT 15
SOR1_MOUSE
ID   SOR1_MOUSE     STANDARD;      PRT;  1167 AA.
AC   Q9JLC4; Q8VI45; Q922I1; Q9QY21;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   VPS10 domain-containing receptor SorCS1 precursor (mSorCS).
GN   SORCS OR SORCS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=C57BL/6J; TISSUE=Brain;
RX   MEDLINE=20068769; PubMed=10600506;
RA   Hermey G., Riedel I.B., Hampe W., Schaller H.C., Hermans-Borgmeyer I.;
RT   "Identification and characterization of SorCS, a third member of a
RT   novel receptor family.";
RL   Biochem. Biophys. Res. Commun. 266:347-351(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=C57BL/6J; TISSUE=Brain;
RX   MEDLINE=20225481; PubMed=10760602;
RA   Hermey G., Schaller H.C.;
RT   "Alternative splicing of murine SorCS leads to two forms of the
RT   receptor that differ completely in their cytoplasmic tails.";
RL   Biochim. Biophys. Acta 1491:350-354(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   STRAIN=C57BL/6J;
RA   Hermey G.;
RT   "A third splice variant of mSorCS.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 424-1167 FROM N.A. (ISOFORM 4).
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=1; Synonyms=SorCSb;
CC         IsoId=Q9JLC4-1; Sequence=Displayed;
CC       Name=2; Synonyms=SorCSa;
CC         IsoId=Q9JLC4-2; Sequence=VSP_006205;
CC       Name=3; Synonyms=SorCSc;
CC         IsoId=Q9JLC4-3; Sequence=VSP_006206;
CC       Name=4;
CC         IsoId=Q9JLC4-4; Sequence=VSP_006207;
CC   -!- TISSUE SPECIFICITY: Isoform 1 is highly expressed in brain, and at
CC       lower levels in heart, liver and kidney. Detected in newborn brain
CC       and in adult olfactory bulb and cerebral cortex. Isoform 2 is
CC       highly expressed in liver, and at lower levels in heart, brain,
CC       kidney and testis.
CC   -!- SIMILARITY: Contains 1 PKD domain.
CC   -!- SIMILARITY: Contains 5 BNR repeats.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to a
CC       frameshift in position 472.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF137367; AAF24748.1; -.
DR   EMBL; AF195056; AAF68196.1; -.
DR   EMBL; AF284755; AAL56666.1; -.
DR   EMBL; BC007486; AAH07486.1; ALT_FRAME.
DR   MGD; MGI:1929666; Sorcs.
DR   InterPro; IPR002860; GH_BNR.
DR   InterPro; IPR000601; PKD.
DR   InterPro; IPR006581; VPS10.
DR   Pfam; PF02012; BNR; 5.
DR   Pfam; PF00801; PKD; 1.
DR   SMART; SM00089; PKD; 2.
DR   SMART; SM00602; VPS10; 1.
DR   PROSITE; PS50093; PKD; 1.
KW   Signal; Transmembrane; Repeat; Alternative splicing.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34   1167       VPS10 DOMAIN-CONTAINING RECEPTOR SORCS1.
FT   DOMAIN       34   1098       LUMENAL (POTENTIAL).
```

```
FT   TRANSMEM    1099   1119        POTENTIAL.
FT   DOMAIN      1120   1167        CYTOPLASMIC (POTENTIAL).
FT   REPEAT       208    219        BNR 1.
FT   REPEAT       256    267        BNR 2.
FT   REPEAT       492    503        BNR 3.
FT   REPEAT       569    580        BNR 4.
FT   REPEAT       611    622        BNR 5.
FT   DOMAIN       802    893        PKD.
FT   CARBOHYD     184    184        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     352    352        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     433    433        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     775    775        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     815    815        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     846    846        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     907    907        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     928    928        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    1124   1167        RVALPSPPSPSAQPGDSSLRLQRPRPATPPSSPKRGSAGAQ
FT                                  FAI -> CVFLLLPSYPRPPPPSSFCQVQKQ (in
FT                                  isoform 2).
FT                                  /FTId=VSP_006205.
FT   VARSPLIC    1124   1167        RVALPSPPSPSAQPGDSSLRLQRPRPATPPSSPKRGSAGAQ
FT                                  FAI -> KIPGINVYAQMQNEKEQELINPVSHSESRPSVPH
FT                                  PDLRRPGQLVDEKVESQLLGK (in isoform 3).
FT                                  /FTId=VSP_006206.
FT   VARSPLIC    1124   1167        RVALPSPPSPSAQPGDSSLRLQRPRPATPPSSPKRGSAGAQ
FT                                  FAI -> KIPGINVYAQMQNEKEQELINPVSHSESRSSVPH
FT                                  PDLRRPGQLVDEKVESQLLGE (in isoform 4).
FT                                  /FTId=VSP_006207.
FT   CONFLICT      10     10        S -> Y (IN REF. 1).
FT   CONFLICT     709    709        Q -> QG (IN REF. 4).
FT   CONFLICT     766    766        A -> T (IN REF. 4).
SQ   SEQUENCE   1167 AA;  129631 MW;  3563396487105B73 CRC64;

  Query Match             71.4%;  Score 40;  DB 1;  Length 1167;
  Best Local Similarity   66.7%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              |:|:||| |
Db       1147 PRPATPPSS 1155


Search completed: April 20, 2004, 10:25:37
Job time : 9.4 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:22:14 ; Search time 10.8 Seconds
(without alignments)
89.066 Million cell updates/sec

Title: US-08-930-480A-7
Perfect score: 56
Sequence: 1 PKPSTPPGSS 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56	100.0	329	1	G3MSC	Ig gamma-3 chain C
2	56	100.0	398	1	G3MSM	Ig gamma-3 chain C
3	43	76.8	209	2	T20975	hypothetical prote
4	43	76.8	341	2	B28820	microtubule-associ
5	43	76.8	364	2	A28820	microtubule-associ
6	43	76.8	374	2	S46264	microtubule-associ
7	43	76.8	432	2	JS0306	microtubule-associ
8	43	76.8	686	2	A38235	microtubule-associ
9	43	76.8	733	2	A45301	microtubule-associ
10	43	76.8	1212	2	S27771	RNA-directed DNA p
11	42	75.0	153	2	F96575	hypothetical prote
12	42	75.0	237	2	G65084	hypothetical prote
13	42	75.0	330	2	D96787	protein T4O12.3 [i
14	42	75.0	418	2	D71460	probable membrane
15	41	73.2	173	2	T31243	hypothetical prote
16	41	73.2	339	2	T46713	hypothetical prote
17	41	73.2	441	2	T12011	cellulase (EC 3.2.
18	40	71.4	122	2	T04366	probable peroxidas
19	40	71.4	190	2	T35570	hypothetical prote
20	40	71.4	229	2	T52364	hypothetical prote
21	40	71.4	235	2	A72594	hypothetical prote
22	40	71.4	274	2	T48819	hypothetical prote
23	40	71.4	463	2	T39004	probable histone a
24	40	71.4	477	2	T46666	N-ethylammeline ch
25	40	71.4	534	2	S21961	proline-rich prote
26	40	71.4	601	2	S56144	SH3 domain binding
27	40	71.4	1014	2	A55260	cytotoxic necrotiz
28	40	71.4	1091	2	S33596	protein-tyrosine k
29	40	71.4	1137	2	A86335	T20H2.9 protein -
30	39.5	70.5	364	1	EDBESP	immediate-early pr
31	39	69.6	154	2	F86306	F20D23.22 protein
32	39	69.6	181	2	T05925	hypothetical prote
33	39	69.6	307	2	T37287	collagen 36 - Caen
34	39	69.6	386	2	I59387	synaptotagmin V -
35	39	69.6	413	2	AH2743	conserved hypothet
36	39	69.6	416	2	G97524	hypothetical prote
37	39	69.6	444	2	C84789	hypothetical prote
38	39	69.6	532	2	JC5412	epidermal growth f
39	39	69.6	536	1	FOMVMD	gag polyprotein -
40	39	69.6	670	2	T49510	fibroin-3 related
41	39	69.6	1014	2	S37405	cytotoxic necrotiz
42	39	69.6	1186	2	T23327	adenomatous polypo
43	39	69.6	1188	2	T23330	hypothetical prote
44	39	69.6	1332	2	T15670	hypothetical prote
45	39	69.6	2517	2	S58380	probable RNA-direc

ALIGNMENTS

RESULT 1
G3MSC
Ig gamma-3 chain C region, secreted form - mouse
C;Species: Mus musculus (house mouse)
C;Date: 17-Mar-1987 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
C;Accession: B02156
R;Wels, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blattr
EMBO J. 3, 2041-2046, 1984
A;Title: Structure analysis of the murine IgG3 constant region gene.
A;Reference number: A02156; MUID:85027161; PMID:6092053
A;Accession: B02156
A;Molecule type: DNA
A;Residues: 1-329 <WEL>
A;Cross-references: GB:J00451
A;Note: the sequence was determined from the germline gene
C;Genetics:
A;Introns: 97/1; 113/1; 223/1
C;Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kapp
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C;Superfamily: immunoglobulin C region; immunoglobulin homology
C;Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobul
F;19-83/Domain: immunoglobulin homology <IM1>
F;97-112/Region: hinge
F;136-205/Domain: immunoglobulin homology <IM2>
F;242-309/Domain: immunoglobulin homology <IM3>
F;179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 56; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 PKPSTPPGSS 10
              ||||||||||
Db        101 PKPSTPPGSS 110
```

RESULT 2
G3MSM
Ig gamma-3 chain C region, membrane-bound form - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Aug-1986 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
C;Accession: A02156; A02155
R;Wels, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blattr
EMBO J. 3, 2041-2046, 1984
A;Title: Structure analysis of the murine IgG3 constant region gene.
A;Reference number: A02156; MUID:85027161; PMID:6092053
A;Accession: A02156
A;Molecule type: DNA
A;Residues: 1-398 <WEL>
A;Cross-references: GB:J00451; NID:g194392; PIDN:AAB59655.1; PID:g194433
A;Note: the sequence was determined from the germline gene
R;Komaromy, M.; Clayton, L.; Rogers, J.; Robertson, S.; Kettman, J.; Wall, R.

```
Nucleic Acids Res. 11, 6775-6785, 1983
A;Title: The structure of the mouse immunoglobulin in gamma-3 membrane gene segment.
A;Reference number: A02155; MUID:84041483; PMID:6314258
A;Accession: A02155
A;Molecule type: DNA
A;Residues: 328-332,'G',334-341,'Q',343-387,'F',389-398 <KOM>
A;Cross-references: GB:K00688
A;Note: the sequence was determined from the germline gene
C;Genetics:
A;Introns: 97/1; 113/1; 223/1; 328/1; 371/3
C;Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C;Superfamily: immunoglobulin C region; immunoglobulin homology
C;Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobu
F;19-83/Domain: immunoglobulin homology <IM1>
F;97-112/Region: hinge
F;136-205/Domain: immunoglobulin homology <IM2>
F;242-309/Domain: immunoglobulin homology <IM3>
F;346-362/Domain: transmembrane #status predicted <TMM>
F;363-398/Domain: intracellular #status predicted <INT>
F;179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             100.0%;  Score 56;  DB 1;  Length 398;
  Best Local Similarity   100.0%;  Pred. No. 0.68;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              ||||||||||
Db        101 PKPSTPPGSS 110


RESULT 3
T20975
hypothetical protein F15D3.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20975
R;White, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19353
A;Accession: T20975
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-209 <WIL>
A;Cross-references: EMBL:Z81063; PIDN:CAB02955.1; GSPDB:GN00019; CESP:F15D3.6
A;Experimental source: clone F15D3
C;Genetics:
A;Gene: CESP:F15D3.6
A;Map position: 1
A;Introns: 11/2; 66/3; 95/3; 153/3; 175/3

  Query Match             76.8%;  Score 43;  DB 2;  Length 209;
  Best Local Similarity   80.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              |||||||  |
Db        199 PKPSTPPPPS 208


RESULT 4
B28820
microtubule-associated protein tau type 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 13-Aug-1999
C;Accession: B28820
R;Lee, G.; Cowan, N.; Kirschner, M.
Science 239, 285-288, 1988
A;Title: The primary structure and heterogeneity of tau protein from mouse brain.
A;Reference number: A94298; MUID:88099510; PMID:3122323
A;Accession: B28820
A;Molecule type: mRNA
A;Residues: 1-341 <LEE>
A;Cross-references: GB:M18775; NID:g201114; PIDN:AAA40165.1; PID:g201115
C;Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C;Keywords: alternative splicing; microtubule binding; tandem repeat
F;183-213/Domain: MAP2/tau repeat homology <MT1>
F;214-244/Domain: MAP2/tau repeat homology <MT2>
F;245-276/Domain: MAP2/tau repeat homology <MT3>

  Query Match             76.8%;  Score 43;  DB 2;  Length 341;
  Best Local Similarity   77.8%;  Pred. No. 43;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | | |||||
Db        108 PSPKTPPGS 116


RESULT 5
A28820
microtubule-associated protein tau type 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 13-Aug-1999
C;Accession: A28820
R;Lee, G.; Cowan, N.; Kirschner, M.
Science 239, 285-288, 1988
A;Title: The primary structure and heterogeneity of tau protein from mouse brain.
A;Reference number: A94298; MUID:88099510; PMID:3122323
A;Accession: A28820
A;Molecule type: mRNA
A;Residues: 1-364 <LEE>
A;Cross-references: GB:M18776; NID:g201116; PIDN:AAA40166.1; PID:g201117
C;Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C;Keywords: alternative splicing; microtubule binding; tandem repeat
F;183-213/Domain: MAP2/tau repeat homology <MT1>
F;214-244/Domain: MAP2/tau repeat homology <MT2>
F;245-276/Domain: MAP2/tau repeat homology <MT3>

  Query Match             76.8%;  Score 43;  DB 2;  Length 364;
  Best Local Similarity   77.8%;  Pred. No. 46;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | | |||||
Db        108 PSPKTPPGS 116


RESULT 6
S46264
microtubule-associated protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 13-Aug-1999
C;Accession: S46264
R;Sadot, E.; Marx, R.; Barg, J.; Behar, L.; Ginzburg, I.
J. Mol. Biol. 241, 325-331, 1994
A;Title: Complete sequence of 3'-untranslated region of tau from rat central nervous syst
A;Reference number: S46264; MUID:94334997; PMID:8057376
A;Accession: S46264
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-374 <SAD>
A;Cross-references: EMBL:X79321; NID:g517393; PIDN:CAA55889.1; PID:g517394
C;Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
F;185-215/Domain: MAP2/tau repeat homology <MT1>
F;216-246/Domain: MAP2/tau repeat homology <MT2>
F;247-277/Domain: MAP2/tau repeat homology <MT3>
F;278-309/Domain: MAP2/tau repeat homology <MT4>

  Query Match             76.8%;  Score 43;  DB 2;  Length 374;
  Best Local Similarity   77.8%;  Pred. No. 47;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              | | |||||
Db        110 PSPKTPPGS 118
```

RESULT 7
JS0306
microtubule-associated protein tau - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 31-Dec-1993
C;Accession: JS0306; A33574
R;Kosik, K.S.; Orecchio, L.D.; Bakalis, S.; Neve, R.L.
Neuron 2, 1389-1397, 1989
A;Title: Developmentally regulated expression of specific tau sequences.
A;Reference number: JS0306; MUID:90180457; PMID:2560640
A;Accession: JS0306
A;Molecule type: mRNA
A;Residues: 1-432 <KOS>
A;Note: the sequence shown is from adult rat brain
A;Note: the partial sequence from fetal rat brain is lacking residues 266-296; the fetal
A;Note: both fetal and adult forms were found in the paired helical filaments characteri
R;Kanai, Y.; Takemura, R.; Oshima, T.; Mori, H.; Ihara, Y.; Yanagisawa, M.; Masaki, T.;
J. Cell Biol. 109, 1173-1184, 1989
A;Title: Expression of multiple tau isoforms and microtubule bundle formation in fibrobl
A;Reference number: A33574; MUID:89359509; PMID:2504728
A;Accession: A33574
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-432 <KAN>
A;Note: a variant lacking residues 63-91 was also found
C;Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C;Keywords: alternative splicing; Alzheimer's disease; calmodulin binding; microtubule b
F;243-273/Domain: MAP2/tau repeat homology <MT1>
F;274-304/Domain: MAP2/tau repeat homology <MT2>
F;305-335/Domain: MAP2/tau repeat homology <MT3>
F;336-367/Domain: MAP2/tau repeat homology <MT4>
F;282-313/Disulfide bonds: #status experimental
F;347/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match 76.8%; Score 43; DB 2; Length 432;
  Best Local Similarity 77.8%; Pred. No. 54;
  Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSTPPGS 9
|| | |||||
Db 168 PSPKTPPGS 176

RESULT 8
A38235
microtubule-associated protein, 110K tau - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Aug-1999
C;Accession: A38235
R;Goedert, M.; Spillantini, M.G.; Crowther, R.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 1983-1987, 1992
A;Title: Cloning of a big tau microtubule-associated protein characteristic of the perip
A;Reference number: A38235; MUID:92179305; PMID:1542696
A;Accession: A38235
A;Molecule type: mRNA
A;Residues: 1-686 <GOE>
A;Cross-references: GB:M84156; NID:g207157; PIDN:AAA42204.1; PID:g207158
A;Note: sequence extracted from NCBI backbone (NCBIN:87358, NCBIP:87359)
C;Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C;Keywords: alternative splicing; microtubule binding; tandem repeat
F;497-527/Domain: MAP2/tau repeat homology <MT1>
F;528-558/Domain: MAP2/tau repeat homology <MT2>
F;559-589/Domain: MAP2/tau repeat homology <MT3>
F;590-621/Domain: MAP2/tau repeat homology <MT4>

  Query Match 76.8%; Score 43; DB 2; Length 686;
  Best Local Similarity 77.8%; Pred. No. 85;
  Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSTPPGS 9
|| | |||||
Db 422 PSPKTPPGS 430

RESULT 9
A45301
microtubule-associated protein tau - mouse
N;Alternate names: microtubule binding protein tau
C;Species: Mus musculus (house mouse)
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 13-Aug-1999
C;Accession: A45301; S31658
R;Couchie, D.; Mavilia, C.; Georgieff, I.S.; Liem, R.K.; Shelanski, M.L.; Nunez, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 4378-4381, 1992
A;Title: Primary structure of high molecular weight tau present in the peripheral nervou
A;Reference number: A45301; MUID:92262443; PMID:1374898
A;Accession: A45301
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-733 <COU>
A;Note: this sequence is inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:102045, NCBIP:102046)
R;Kenner, L.; Forstner, M.; Hutter, H.; Hoefler, G.; Kurzbauer, R.; Zatloukal, K.; Krisp
submitted to the EMBL Data Library, May 1992
A;Description: First observation of mRNA for a tau-protein from murine liver and kidney.
A;Reference number: S31658
A;Accession: S31658
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 'T',529-651 <KEN>
A;Cross-references: EMBL:Z12133; NID:g54263; PIDN:CAA78121.1; PID:g388534
C;Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C;Keywords: microtubule binding; tandem repeat
F;544-574/Domain: MAP2/tau repeat homology <MT1>
F;575-605/Domain: MAP2/tau repeat homology <MT2>
F;606-636/Domain: MAP2/tau repeat homology <MT3>
F;637-668/Domain: MAP2/tau repeat homology <MT4>

  Query Match 76.8%; Score 43; DB 2; Length 733;
  Best Local Similarity 77.8%; Pred. No. 91;
  Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSTPPGS 9
|| | |||||
Db 469 PSPKTPPGS 477

RESULT 10
S27771
RNA-directed DNA polymerase (EC 2.7.7.49) - African malaria mosquito transposon RT1 (fra
N;Alternate names: reverse transcriptase
C;Species: Anopheles gambiae (African malaria mosquito)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C;Accession: S27771
R;Besansky, N.J.; Paskewitz, S.M.; Mills-Hamm, D.M.; Collins, F.H.
submitted to the EMBL Data Library, June 1992
A;Description: Distinct families of site-specific retroposons occupy identical positions
A;Reference number: S27770
A;Accession: S27771
A;Molecule type: DNA
A;Residues: 1-1212 <BES>
A;Cross-references: EMBL:M93690; NID:g159615; PID:g159617
C;Keywords: nucleotidyltransferase

  Query Match 76.8%; Score 43; DB 2; Length 1212;
  Best Local Similarity 70.0%; Pred. No. 1.5e+02;
  Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||| |||| :
Db 545 PKPGKPPGSN 554

RESULT 11
F96575

```
hypothetical protein F22G10.10 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: F96575
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F96575
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <STO>
A;Cross-references: GB:AE005173; NID:g10645344; PIDN:AAG21464.1; GSPDB:GN00141
C;Genetics:
A;Gene: F22G10.10
A;Map position: 1

  Query Match             75.0%;  Score 42;  DB 2;  Length 153;
  Best Local Similarity   77.8%;  Pred. No. 28;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGS 9
              |:|| ||||
Db         43 PQPSPPPGS 51


RESULT 12
G65084
hypothetical protein b2985 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: G65084
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G65084
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-237 <BLAT>
A;Cross-references: GB:AE000381; GB:U00096; NID:g2367181; PIDN:AAC76021.1; PID:g1789358;
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: conserved hypothetical protein b2986

  Query Match             75.0%;  Score 42;  DB 2;  Length 237;
  Best Local Similarity   87.5%;  Pred. No. 42;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KPSTPPGS 9
              |||||||:
Db         84 KPSTPPGN 91


RESULT 13
D96787
protein T4O12.3 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 24-Aug-2001
C;Accession: D96787
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-330 <STO>
A;Cross-references: GB:AE005173; NID:g6721098; PIDN:AAF26752.1; GSPDB:GN00141
C;Genetics:
A;Gene: T4O12.3
A;Map position: 1
C;Superfamily: thaumatin I

  Query Match             75.0%;  Score 42;  DB 2;  Length 330;
  Best Local Similarity   70.0%;  Pred. No. 58;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              |||:|| |:|
Db        264 PKPTTPTGTS 273


RESULT 14
D71460
probable membrane thiol proteinase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: D71460
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: D71460
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-418 <ARN>
A;Cross-references: GB:AE001360; GB:AE001273; NID:g3329342; PIDN:AAC68466.1; PID:g332934
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: CT868

  Query Match             75.0%;  Score 42;  DB 2;  Length 418;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PKPSTPP 7
              |||||||
Db        128 PKPSTPP 134


RESULT 15
T31243
hypothetical protein 915 - Sphingomonas aromaticivorans plasmid pNL1
C;Species: Sphingomonas aromaticivorans
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C;Accession: T31243
R;Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.; G
submitted to the EMBL Data Library, July 1998
A;Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromati
A;Reference number: Z20992
A;Accession: T31243
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-173 <ROM>
A;Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378383; PIDN:AAD03966.1
C;Genetics:
A;Genome: plasmid pNL1
A;Note: orf915
C;Superfamily: Sphingomonas aromaticivorans hypothetical protein 915

  Query Match             73.2%;  Score 41;  DB 2;  Length 173;
```

Best Local Similarity 87.5%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 KPSTPPGS 9
              |||:||||
Db         13 KPSSPPGS 20
```

Search completed: April 20, 2004, 10:27:42
Job time : 11.8 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:20:04 ; Search time 40.4 Seconds
(without alignments)
69.938 Million cell updates/sec

Title: US-08-930-480A-7
Perfect score: 56
Sequence: 1 PKPSTPPGSS 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56	100.0	10	2	AAW09324	Aaw09324 Peptide l
2	56	100.0	10	5	ABG97610	Abg97610 Apolipopr
3	56	100.0	16	2	AAW71022	Aaw71022 Mus muscu
4	56	100.0	17	5	ABG94379	Abg94379 N termina
5	56	100.0	17	5	ABG80675	Abg80675 N-termina
6	56	100.0	17	6	ABR56408	Abr56408 Peptide l
7	56	100.0	17	6	ABR44509	Abr44509 Peptide l
8	56	100.0	17	7	ADD24209	Add24209 Linker pe
9	56	100.0	18	5	ABG94380	Abg94380 C termina
10	56	100.0	18	5	ABG80676	Abg80676 C-termina
11	56	100.0	18	6	ABR56409	Abr56409 Peptide l
12	56	100.0	18	6	ABR44510	Abr44510 Peptide l
13	56	100.0	18	7	ADD24210	Add24210 Linker pe
14	56	100.0	42	3	AAY91034	Aay91034 Lipid-tag
15	56	100.0	46	2	AAR39337	Aar39337 Intercala
16	56	100.0	47	2	AAW22021	Aaw22021 Di-alpha-
17	56	100.0	50	2	AAR39338	Aar39338 Intercala
18	56	100.0	53	2	AAR39340	Aar39340 Intercala
19	56	100.0	53	2	AAR39339	Aar39339 Intercala
20	56	100.0	81	5	ABP51694	Abp51694 Plasmid p
21	56	100.0	81	5	ABB79463	Abb79463 Recombina
22	56	100.0	134	5	ABG94335	Abg94335 Human MIF
23	56	100.0	134	5	ABG94336	Abg94336 met-human
24	56	100.0	134	5	ABG80648	Abg80648 Human MIF
25	56	100.0	135	5	ABG94330	Abg94330 rMIF-C3 p
26	56	100.0	135	5	ABG94334	Abg94334 Human MIF
27	56	100.0	135	5	ABG80642	Abg80642 Rat MIF,
28	56	100.0	135	5	ABG80647	Abg80647 Human MIF
29	56	100.0	136	5	ABG94348	Abg94348 Mouse C-I
30	56	100.0	136	5	ABG80660	Abg80660 Mouse IL-
31	56	100.0	138	5	ABG94351	Abg94351 Human C-I
32	56	100.0	138	5	ABG80663	Abg80663 Human ILn
33	56	100.0	263	5	ABG80714	Abg80714 Human IgG
34	56	100.0	322	4	AAB20440	Aab20440 Antibody
35	56	100.0	325	4	AAB20438	Aab20438 Anti-FIX/
36	56	100.0	450	3	AAY44991	Aay44991 M79scFv-i
37	56	100.0	466	6	ABU62399	Abu62399 Chimeric
38	56	100.0	467	6	AAE38408	Aae38408 Mouse vir
39	56	100.0	531	2	AAR98007	Aar98007 PelB sign
40	45	80.4	782	4	AAO00740	Aao00740 Human pol
41	43	76.8	116	4	AAU48146	Aau48146 Propionib
42	43	76.8	116	6	ABM44665	Abm44665 Propionib
43	43	76.8	364	5	ABB57300	Abb57300 Mouse isc
44	43	76.8	374	6	ABM04837	Abm04837 Rat tau m
45	43	76.8	446	4	ABG06313	Abg06313 Novel hum

ALIGNMENTS

RESULT 1
AAW09324
ID AAW09324 standard; peptide; 10 AA.
XX
AC AAW09324;
XX
DT 10-JUN-1997 (first entry)
XX
DE Peptide linker arm #2.
XX
KW Chimaeric; bispecific; DNA binding domain; trans; activator; repressor;
KW diphtheria; Pseudomonas; toxin; thymidine kinase; single chain antibody;
KW pathogen; HIV Tat; papilloma virus; E6/E7; Epstein-Barr virus; EBNA;
KW hyperproliferation; p53; tumour; oligomerisation.
XX
OS Synthetic.
XX
PN WO9630512-A1.
XX
PD 03-OCT-1996.
XX
PF 29-MAR-1996; 96WO-FR000477.
XX
PR 31-MAR-1995; 95FR-00003841.
XX
PA (RHON) RHONE POULENC RORER SA.
XX
PI Bracco L, Schweighoffer F, Tocque B;
XX
DR WPI; 1996-455359/45.
DR N-PSDB; AAT47997.
XX
PT Conditional gene expression system triggered by e.g. infection or hyper-
PT proliferation - comprises novel bi:specific proteins having DNA-binding
PT domain and second domain specific for trans-activator or repressor, for
PT gene therapy.
XX
PS Claim 23; Page 46; 81pp; French.
XX
CC The invention relates to novel chimaeric, bispecific proteins which
CC comprise: (a) a DNA binding domain and (b) a domain which binds a trans-
CC activator (TA), trans-repressor (TR) or their complexes, which are
CC characteristic of a physiological or physiopathological state. The novel
CC chimaeric, bispecific proteins allow expression of a therapeutic protein
CC (e.g. diphtheria or Pseudomonas toxins, thymidine kinase, single chain
CC antibodies) to be regulated in response to particular conditions.
CC Examples include making the protein responsive to the presence of
CC particular pathogenic TA mols (e.g. HIV Tat, papilloma virus E6/E7

CC proteins or Epstein-Barr virus EBNA protein), the therapeutic protein
CC will be expressed in those cells infected by that pathogen. Similarly,
CC where the chimaeric protein responds to a cellular protein typical of a
CC hyperproliferative state (esp. wild-type and mutant p53), expression can
CC be restricted to tumour cells. The sequence presented here is an example
CC of a peptide linker "arm" which connects the DNA binding domain to the TA
CC binding domain
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 56; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 1 PKPSTPPGSS 10

RESULT 2
ABG97610
ID ABG97610 standard; peptide; 10 AA.
XX
AC ABG97610;
XX
DT 17-DEC-2002 (first entry)
XX
DE Apolipoprotein analogue 1 (Apo A1) associated spacer peptide #4.
XX
KW Apolipoprotein analogue; Apo A; arteriosclerosis; endotoxin removal;
KW angina pectoris; myocardial infarction; arterial stenosis; claudicatio;
KW carotis stenosis; cerebral arterial stenosis; gene therapy; cholesterol;
KW cardiovascular disease; spacer peptide.
XX
OS Mus sp.
XX
PN WO200238609-A2.
XX
PD 16-MAY-2002.
XX
PF 09-NOV-2001; 2001WO-DK000739.
XX
PR 10-NOV-2000; 2000DK-00001682.
PR 15-JAN-2001; 2001DK-00000057.
PR 26-JAN-2001; 2001US-0264022P.
XX
PA (PROT-) PROTEOPHARMA APS.
XX
PI Graversen J, Moestrup S;
XX
DR WPI; 2002-527481/56.
XX
PT Novel apolipoprotein construct comprising apolipoprotein A linked to
PT carbohydrate, peptide or protein heterologous group, useful for treating
PT plaque/unstable angina pectoris, myocardial infarction, arterial
PT stenoses.
XX
PS Claim 8; Page 54; 113pp; English.
XX
CC The invention describes an Apolipoprotein (Apo) construct (I) for use as
CC medicament having general formula apo-A-X, where apo-A is an
CC apolipoprotein component such as apolipoprotein AI, AII or AIV, or its
CC analogue or variant, and X is heterologous group e.g., amino acid,
CC peptide, protein, carbohydrate or a nucleic acid, providing that when (I)
CC consists of exactly two identical, native apolipoproteins these are
CC linked serially. (I) is useful for preparing a pharmaceutical composition
CC which further comprises excipients, adjuvants, additives, such as
CC phospholipids, cholesterol or triglycerides. (I) is useful or treating
CC and/or preventing arteriosclerosis, for removing endotoxins, for treating
CC angina pectoris including plaque or unstable angina pectoris, myocardial
CC infarction, arterial stenoses such as claudicatio, carotis stenosis,
CC cerebral arterial stenosis and other cardiovascular diseases. The nucleic
CC acid (II) encoding (I) is useful for gene therapy, where the DNA sequence
CC encoding (I) is used for transfection or infection of at least one cell
CC population comprising macrophages or liver cells. (I) has a half-life of
CC at least the half-life of native apoA-I, A-II or A-IV, preferably two
CC times higher or more preferably 10 times higher than the half-life of the
CC apoA molecules. (I) also has a higher binding affinity to cholesterol
CC compared to native apoA-I, A-II or A-IV. (I) causes substantially no
CC immune response in humans. This is the amino acid sequence of a spacer
CC peptide used to link human apolipoprotein (Apo) or an Apo analogue
CC protein to a heterologous moiety
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 56; DB 5; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 1 PKPSTPPGSS 10

RESULT 3
AAW71022
ID AAW71022 standard; protein; 16 AA.
XX
AC AAW71022;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mus musculus Cgamma3 gene hinge region.
XX
KW Cgamma3 gene; transgenic; screening; therapeutic; nephritis;
KW systemic lupus erythematosus.
XX
OS Mus musculus.
XX
PN WO9837174-A1.
XX
PD 27-AUG-1998.
XX
PF 18-FEB-1998; 98WO-US003027.
XX
PR 20-FEB-1997; 97US-00803120.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI Schreiber JR, Greenspan NS, Threadgill DS, Magnuson T;
XX
DR WPI; 1998-467546/40.
DR N-PSDB; AAV43042.
XX
PT New transgenic animals lacking IgG3 - is used for screening candidate
PT therapeutic compounds, in particular for activity against bacterial
PT infection or nephritis.
XX
PS Disclosure; Fig 1; 60pp; English.
XX
CC The sequence is that of the hinge region of the protein encoded by the
CC mouse Cgamma3 gene
XX
SQ Sequence 16 AA;

Query Match 100.0%; Score 56; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 4 PKPSTPPGSS 13

RESULT 4
ABG94379

ID ABG94379 standard; peptide; 17 AA.
XX
AC ABG94379;
XX
DT 10-DEC-2002 (first entry)
XX
DE N terminal gamma 3 amino acid linker.
XX
KW Human; mouse; rat; antimicrobial; antiallergic; immunomodulatory;
KW cytostatic; antiviral; antidiabetic; hypoglycaemic; antigen array;
KW vaccine; infectious disease.
XX
OS Synthetic.
XX
PN WO200256905-A2.
XX
PD 25-JUL-2002.
XX
PF 21-JAN-2002; 2002WO-IB000166.
XX
PR 19-JAN-2001; 2001US-0262379P.
PR 04-MAY-2001; 2001US-0288549P.
PR 05-OCT-2001; 2001US-0326998P.
PR 07-NOV-2001; 2001US-0331045P.
XX
PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
XX
PI Renner WA, Bachmann M, Tissot A, Maurer P, Lechner F, Sebbel P;
PI Piossek C;
XX
DR WPI; 2002-627351/67.
XX
PT Molecular antigen array used in the production of vaccines for infectious
PT diseases.
XX
PS Disclosure; Page 49; 441pp; English.
XX
CC This invention relates to a novel ordered and repetitive antigen array
CC used in the production of vaccines for infectious diseases. The invention
CC also discloses a composition comprising a non-natural molecular scaffold
CC comprising a core particle selected from a core particle of a non-natural
CC origin and a core particle of natural origin and an organiser comprising
CC at least one first attachment site, where the organiser is connected to
CC the core particle by at least one covalent bond. Also disclosed is an
CC antigen or antigenic determinant with at least one second attachment
CC site, where the antigen or antigenic determinant is amyloid beta peptide
CC (Abeta1-42) or its fragment and where the second attachment site is
CC selected from an attachment site not naturally occurring with the antigen
CC or antigenic determinant and an attachment site naturally occurring with
CC the antigen or antigenic determinant, where the second attachment site is
CC capable of association through at least one non-peptide bond to the first
CC attachment site and where the antigen or antigenic determinant and the
CC scaffold interact through the association to form an ordered and
CC repetitive antigen array. The invention also comprises a coat protein
CC capable of forming a capsid which comprises mutant Qbeta coat proteins
CC having an amino acid sequence selected from five amino acid sequences
CC fully defined in the specification. The compounds of the invention may
CC have antimicrobial, antiallergic, immunomodulatory, cytostatic,
CC antiviral, antidiabetic, or hypoglycaemic activities and may be used in
CC immunization and as a vaccine. The present sequence represents a peptide
CC sequence used to create the compositions of the invention
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 56; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 4 PKPSTPPGSS 13

RESULT 5
ABG80675
ID ABG80675 standard; peptide; 17 AA.
XX
AC ABG80675;
XX
DT 29-NOV-2002 (first entry)
XX
DE N-terminal gamma 3 linker peptide.
XX
KW Molecular antigen array; vaccine; antigen; antimicrobial;
KW molecular scaffold; amyloid beta; Abeta 1-42; influenza;
KW graft versus host disease; IgE-mediated allergic reaction; anaphylaxis;
KW adult respiratory distress syndrome; ARDS; Crohn's disease;
KW allergic asthma; acute lymphoblastic leukaemia; non-Hodgkin's lymphoma;
KW Grave's disease; systemic lupus erythematosus; osteoporosis;
KW inflammatory immune disease; myasthenia gravis; multiple sclerosis;
KW immunoproliferative disease lymphadenopathy; Alzheimer's disease;
KW angioimmunoproliferative lymphadenopathy; immunoblastive lymphadenopathy;
KW rheumatoid arthritis; diabetes; infectious disease; factor Xa;
KW enterokinase; cysteine-containing linker.
XX
OS Synthetic.
XX
PN WO200256907-A2.
XX
PD 25-JUL-2002.
XX
PF 21-JAN-2002; 2002WO-IB000168.
XX
PR 19-JAN-2001; 2001US-0262379P.
PR 04-MAY-2001; 2001US-0288549P.
PR 05-OCT-2001; 2001US-0326998P.
PR 07-NOV-2001; 2001US-0331045P.
XX
PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
PA (NOVS) NOVARTIS PHARMA AG.
PA (MAUR/) MAURER P.
PA (LECH/) LECHNER F.
PA (ORTM/) ORTMANN R.
PA (LUEO/) LUEOEND R.
PA (STAU/) STAUFENBIEL M.
PA (FREY/) FREY P.
XX
PI Maurer P, Lechner F, Ortmann R, Lueoend R, Staufenbiel M, Frey P;
PI Renner WA, Bachmann M, Tissot A, Sebbel P, Piossek C;
XX
DR WPI; 2002-636514/68.
XX
PT Molecular antigen array used in the production of vaccines for infectious
PT diseases.
XX
PS Claim 35; Page 49; 418pp; English.
XX
CC The invention relates to a composition comprising: (a) a non-natural
CC molecular scaffold comprising: (i) a core particle selected from: (1) a
CC core particle of a non-natural origin; and (2) a core particle of natural
CC origin; and (ii) an organiser comprising at least one first attachment
CC site, where the organiser is connected to the core particle by at least
CC one covalent bond; (b) an antigen or antigenic determinant with at least
CC one second attachment site, where the antigen or antigenic determinant is
CC amyloid beta peptide (Abeta 1-42) or its fragment, and where the second
CC attachment site is selected from: (i) an attachment site not naturally
CC occurring with the antigen or antigenic determinant; and (ii) an
CC attachment site naturally occurring with the antigen or antigenic
CC determinant, where the second attachment site is capable of association
CC through at least one non-peptide bond to the first attachment site; and
CC where the antigen or antigenic determinant and the scaffold interact
CC through the association to form an ordered and repetitive antigen array.
CC Also included is a process for producing a non-naturally occurring
CC ordered and repetitive antigen array. The composition is used in
CC immunisation and as a vaccine for diseases such as influenza, graft
CC versus host disease, IgE-mediated allergic reactions, anaphylaxis, adult

```
CC   respiratory distress syndrome (ARDS), Crohn's disease, allergic asthma,
CC   acute lymphoblastic leukaemia, non-Hodgkin's lymphoma, Grave's disease,
CC   systemic lupus erythematosus, inflammatory immune diseases, myasthenia
CC   gravis, immunoproliferative disease lymphadenopathy,
CC   angioimmunoproliferative lymphadenopathy, immunoblastive lymphadenopathy,
CC   rheumatoid arthritis, diabetes, multiple sclerosis, Alzheimer's disease,
CC   osteoporosis and infectious diseases. The antigens are modified to posses
CC   a cleavage site (enterokinase or factor Xa) and a Cysteine- containing N-
CC   or C-terminal linker peptide which serves as the attachment point to a
CC   virus like particle or bacterial protein (the scaffold protein). The
CC   present sequence is a cysteine-containing linker peptide used in the
CC   molecular antigen array
XX
SQ   Sequence 17 AA;

  Query Match             100.0%;  Score 56;  DB 5;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 1.4;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              ||||||||||
Db          4 PKPSTPPGSS 13


RESULT 6
ABR56408
ID   ABR56408 standard; peptide; 17 AA.
XX
AC   ABR56408;
XX
DT   28-JUL-2003  (first entry)
XX
DE   Peptide linker #8.
XX
KW   Antigen presenting cell; APC; immune response; virus like particle; VLP;
KW   cytostatic; virucide; antibacterial; antiparasitic; fungicide;
KW   antiallergic; immunosuppressive; antiaddictive; antiinflammatory;
KW   antithyroid; antidiabetic; neuroprotective; nootropic; osteopathic;
KW   antirheumatic; antiarthritic; vaccine; immunisation; infectious disease;
KW   anti-viral protection; tumour; allergy; drug addiction; Crohn's disease;
KW   graft-versus-host disease; Grave's disease; diabetes; multiple sclerosis;
KW   Alzheimer's disease; osteoporosis; rheumatoid arthritis;
KW   inflammatory autoimmune disease.
XX
OS   Synthetic.
XX
PN   WO2003024480-A2.
XX
PD   27-MAR-2003.
XX
PF   16-SEP-2002; 2002WO-IB004252.
XX
PR   14-SEP-2001; 2001US-0318967P.
XX
PA   (CYTO-) CYTOS BIOTECHNOLOGY AG.
XX
PI   Bachmann MF,  Storni T,  Lechner F;
XX
DR   WPI; 2003-363095/34.
XX
PT   A composition, useful for enhancing an immune response against an antigen
PT   or a virus-like particle, enhancing anti-viral protection in an animal,
PT   or immunizing or treating tumors or infectious diseases, e.g. viral
PT   infections.
XX
PS   Disclosure; Page 65; 243pp; English.
XX
CC   The present invention describes a composition (C) for enhancing an immune
CC   response against an antigen or a virus-like particle in an animal. (C)
CC   comprises a virus-like particle (VLP) bound to at least one antigen, or a
CC   VLP capable of being recognised by the immune system of the animal. Also
CC   described: (1) enhancing an immune response against an antigen or a VLP
CC   in an animal comprising introducing (C) into the animal; (2) vaccines
CC   comprising (C) together with a pharmaceutical diluent, carrier or
CC   excipient; (3) immunising or treating an animal comprising administering
CC   the vaccine to the animal, or priming or boosting a T cell response in
CC   the animal by administering the vaccine; and (4) enhancing anti-viral
CC   protection in an animal comprising introducing (C) into the animal. (C)
CC   has cytostatic, virucide, antibacterial, antiparasitic, fungicide,
CC   antiallergic, immunosuppressive, antiaddictive, antiinflammatory,
CC   antithyroid, antidiabetic, neuroprotective, nootropic, osteopathic,
CC   antirheumatic and antiarthritic activities. (C) or the vaccines can be
CC   used for enhancing an immune response against an antigen or a VLP in an
CC   animal, enhancing anti-viral protection in an animal, or immunising or
CC   treating tumours and infectious diseases such as viral, bacterial,
CC   parasitic or fungal infections. The vaccine compositions are also useful
CC   for preventing or treating allergies, drug addiction, graft-versus-host
CC   disease, Crohn's disease, Grave's disease, diabetes, multiple sclerosis,
CC   Alzheimer's disease, osteoporosis, rheumatoid arthritis, or inflammatory
CC   autoimmune disease. ACC69838 to ACC69852 and ABR56401 to ABR56509
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 17 AA;

  Query Match             100.0%;  Score 56;  DB 6;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 1.4;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              ||||||||||
Db          4 PKPSTPPGSS 13


RESULT 7
ABR44509
ID   ABR44509 standard; peptide; 17 AA.
XX
AC   ABR44509;
XX
DT   25-JUL-2003  (first entry)
XX
DE   Peptide linker #8.
XX
KW   Immunostimulatory; virus-like particle; bacteriophage; HBV; LCMV;
KW   hepatitis B virus; lymphocytic choriomeningitis virus; vaccine;
KW   immunostimulant; cytostatic; antiallergic; virucide; antibacterial;
KW   immune response; immunisation; allergy; tumour; breast cancer;
KW   neuroblastoma; leukaemia; viral disease; influenza; hepatitis; measles;
KW   chicken pox; bacterial infection; tuberculosis; pneumonia; syphilis.
XX
OS   Synthetic.
XX
PN   WO2003024481-A2.
XX
PD   27-MAR-2003.
XX
PF   16-SEP-2002; 2002WO-IB004132.
XX
PR   14-SEP-2001; 2001US-0318994P.
PR   22-APR-2002; 2002US-0374145P.
XX
PA   (CYTO-) CYTOS BIOTECHNOLOGY AG.
PA   (MAUR/) MAURER P.
PA   (TISS/) TISSOT A.
PA   (SCHW/) SCHWARZ K.
PA   (MEIJ/) MEIJERINK E.
PA   (LIPO/) LIPOWSKY G.
PA   (PUMP/) PUMPENS P.
PA   (CIEL/) CIELENS I.
PA   (RENH/) RENHOFA R.
XX
PI   Maurer P,  Tissot A,  Schwarz K,  Meijerink E,  Lipowsky G;
PI   Pumpens P,  Cielens I,  Renhofa R,  Bachmann MF,  Storni T;
XX
DR   WPI; 2003-354564/33.
XX
```

PT New compositions comprising immunostimulatory substances packaged into
PT virus-like particles, useful as a vaccine for enhancing an immune
PT response in animals, e.g. for treating or preventing allergies, tumors or
PT viral infections.
XX
PS Disclosure; Page 75; 322pp; English.
XX
CC The present invention describes a composition (C) for enhancing an immune
CC response in an animal. (C) comprises a virus-like particle (VLP), and an
CC immunostimulatory substance. The immunostimulatory substance is bound to
CC the VLP. Also described: (1) enhancing an immune response in an animal by
CC introducing (C) into the animal; (2) producing (C) for enhancing an
CC immune response in an animal; (3) vaccines comprising (C) together with a
CC pharmaceutical diluent, carrier or excipient; and (4) immunising or
CC treating an animal by: (a) administering the vaccine to the animal; (b)
CC priming a T cell response in the animal by administering the vaccine; or
CC (c) boosting a T cell response in the animal by administering the
CC vaccine. (C) has immunostimulant, cytostatic, antiallergic, virucide and
CC antibacterial activities. (I) can be used in vaccines for enhancing an
CC immune response in an animal, particularly a mammal or human.
CC Specifically, (C) is useful for enhancing a B cell response, a T cell
CC response, or a cytotoxic T-lymphocyte (CTL) response. (C) or a vaccine
CC comprising (C) can also be used for immunising or treating an animal,
CC e.g. humans, sheep, horses, cattle, pigs, dogs, cats, rats, birds,
CC reptiles or fish. (C) is particularly useful in prophylactic or
CC therapeutic vaccines against allergies, tumours (e.g. breast cancers,
CC neuroblastoma, or leukaemia), viral diseases (e.g. influenza, hepatitis,
CC measles or chicken pox), or bacterial infections (e.g. tuberculosis,
CC pneumonia or syphilis). ACC69790 to ACC69815 and ABR44502 to ABR44612
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 56; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 4 PKPSTPPGSS 13

RESULT 8
ADD24209
ID ADD24209 standard; peptide; 17 AA.
XX
AC ADD24209;
XX
DT 15-JAN-2004 (first entry)
XX
DE Linker peptide 3 related to prion disease vaccines.
XX
KW vaccine composition; virus-like particle; core particle;
KW first attachment site; antigen; antigenic determinant; prion protein;
KW PrP; PrP peptide; vaccine; neuroprotective; antiinflammatory;
KW prion disease; Bovine Spongiform Encephalopathy; BSE;
KW Creutzfeldt-Jakob Disease; linker peptide.
XX
OS Synthetic.
XX
PN WO2003059386-A2.
XX
PD 24-JUL-2003.
XX
PF 17-JAN-2003; 2003WO-EP000460.
XX
PR 18-JAN-2002; 2002US-00050902.
PR 21-JAN-2002; 2002WO-IB000166.
PR 08-JUL-2002; 2002US-0393725P.
PR 18-JUL-2002; 2002US-0396590P.
XX
PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
XX
PI Bachmann M, Maurer P, Pellicioli E, Renner WA;
XX
DR WPI; 2003-598483/56.
XX
PT A vaccine composition for preventing or treating prion diseases (e.g.
PT Creutzfeldt-Jakob Disease) comprises a virus-like particle (e.g. RNA-
PT phage) and at least one prion protein or peptide bound to the virus-like
PT particle.
XX
PS Disclosure; Page 59; 246pp; English.
XX
CC This invention relates to a novel vaccine composition comprising a virus-
CC like or a core particle with at least one first attachment site and at
CC least one antigen or antigenic determinant that is a prion protein (PrP)
CC or its dimer, or a PrP peptide, the antigen or antigenic determinant
CC being bound to the virus-like or core particle. The vaccine of the
CC invention may have neuroprotective or antiinflammatory activity. The
CC composition is useful as a medicament or in manufacturing a medicament
CC for the treatment or prevention of prion diseases. The prion diseases may
CC include Bovine Spongiform Encephalopathy (BSE) or Creutzfeldt-Jakob
CC Disease. The present sequence is that of a linker peptide which is
CC related to the invention.
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 56; DB 7; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 4 PKPSTPPGSS 13

RESULT 9
ABG94380
ID ABG94380 standard; peptide; 18 AA.
XX
AC ABG94380;
XX
DT 10-DEC-2002 (first entry)
XX
DE C terminal gamma 3 amino acid linker.
XX
KW Human; mouse; rat; antimicrobial; antiallergic; immunomodulatory;
KW cytostatic; antiviral; antidiabetic; hypoglycaemic; antigen array;
KW vaccine; infectious disease.
XX
OS Synthetic.
XX
PN WO200256905-A2.
XX
PD 25-JUL-2002.
XX
PF 21-JAN-2002; 2002WO-IB000166.
XX
PR 19-JAN-2001; 2001US-0262379P.
PR 04-MAY-2001; 2001US-0288549P.
PR 05-OCT-2001; 2001US-0326998P.
PR 07-NOV-2001; 2001US-0331045P.
XX
PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
XX
PI Renner WA, Bachmann M, Tissot A, Maurer P, Lechner F, Sebbel P;
PI Piossek C;
XX
DR WPI; 2002-627351/67.
XX
PT Molecular antigen array used in the production of vaccines for infectious
PT diseases.
XX
PS Disclosure; Page 49; 441pp; English.
XX

CC This invention relates to a novel ordered and repetitive antigen array
CC used in the production of vaccines for infectious diseases. The invention
CC also discloses a composition comprising a non-natural molecular scaffold
CC comprising a core particle selected from a core particle of a non-natural
CC origin and a core particle of natural origin and an organiser comprising
CC at least one first attachment site, where the organiser is connected to
CC the core particle by at least one covalent bond. Also disclosed is an
CC antigen or antigenic determinant with at least one second attachment
CC site, where the antigen or antigenic determinant is amyloid beta peptide
CC (Abeta1-42) or its fragment and where the second attachment site is
CC selected from an attachment site not naturally occurring with the antigen
CC or antigenic determinant and an attachment site naturally occurring with
CC the antigen or antigenic determinant, where the second attachment site is
CC capable of association through at least one non-peptide bond to the first
CC attachment site and where the antigen or antigenic determinant and the
CC scaffold interact through the association to form an ordered and
CC repetitive antigen array. The invention also comprises a coat protein
CC capable of forming a capsid which comprises mutant Qbeta coat proteins
CC having an amino acid sequence selected from five amino acid sequences
CC fully defined in the specification. The compounds of the invention may
CC have antimicrobial, antiallergic, immunomodulatory, cytostatic,
CC antiviral, antidiabetic, or hypoglycaemic activities and may be used in
CC immunization and as a vaccine. The present sequence represents a peptide
CC sequence used to create the compositions of the invention
XX
SQ Sequence 18 AA;

Query Match 100.0%; Score 56; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 1 PKPSTPPGSS 10

RESULT 10
ABG80676
ID ABG80676 standard; peptide; 18 AA.
XX
AC ABG80676;
XX
DT 29-NOV-2002 (first entry)
XX
DE C-terminal gamma 3 linker peptide.
XX
KW Molecular antigen array; vaccine; antigen; antimicrobial;
KW molecular scaffold; amyloid beta; Abeta 1-42; influenza;
KW graft versus host disease; IgE-mediated allergic reaction; anaphylaxis;
KW adult respiratory distress syndrome; ARDS; Crohn's disease;
KW allergic asthma; acute lymphoblastic leukaemia; non-Hodgkin's lymphoma;
KW Grave's disease; systemic lupus erythematosus; osteoporosis;
KW inflammatory immune disease; myasthenia gravis; multiple sclerosis;
KW immunoproliferative disease lymphadenopathy; Alzheimer's disease;
KW angioimmunoproliferative lymphadenopathy; immunoblastive lymphadenopathy;
KW rheumatoid arthritis; diabetes; infectious disease; factor Xa;
KW enterokinase; cysteine-containing linker.
XX
OS Synthetic.
XX
PN WO200256907-A2.
XX
PD 25-JUL-2002.
XX
PF 21-JAN-2002; 2002WO-IB000168.
XX
PR 19-JAN-2001; 2001US-0262379P.
PR 04-MAY-2001; 2001US-0288549P.
PR 05-OCT-2001; 2001US-0326998P.
PR 07-NOV-2001; 2001US-0331045P.
XX
PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
PA (NOVS) NOVARTIS PHARMA AG.
PA (MAUR/) MAURER P.
PA (LECH/) LECHNER F.
PA (ORTM/) ORTMANN R.
PA (LUEO/) LUEOEND R.
PA (STAU/) STAUFENBIEL M.
PA (FREY/) FREY P.
XX
PI Maurer P, Lechner F, Ortmann R, Lueoend R, Staufenbiel M, Frey P;
PI Renner WA, Bachmann M, Tissot A, Sebbel P, Piossek C;
XX
DR WPI; 2002-636514/68.
XX
PT Molecular antigen array used in the production of vaccines for infectious
PT diseases.
XX
PS Claim 35; Page 49; 418pp; English.
XX
CC The invention relates to a composition comprising: (a) a non-natural
CC molecular scaffold comprising: (i) a core particle selected from: (1) a
CC core particle of a non-natural origin; and (2) a core particle of natural
CC origin; and (ii) an organiser comprising at least one first attachment
CC site, where the organiser is connected to the core particle by at least
CC one covalent bond; (b) an antigen or antigenic determinant with at least
CC one second attachment site, where the antigen or antigenic determinant is
CC amyloid beta peptide (Abeta 1-42) or its fragment, and where the second
CC attachment site is selected from: (i) an attachment site not naturally
CC occurring with the antigen or antigenic determinant; and (ii) an
CC attachment site naturally occurring with the antigen or antigenic
CC determinant, where the second attachment site is capable of association
CC through at least one non-peptide bond to the first attachment site; and
CC where the antigen or antigenic determinant and the scaffold interact
CC through the association to form an ordered and repetitive antigen array.
CC Also included is a process for producing a non-naturally occurring
CC ordered and repetitive antigen array. The composition is used in
CC immunisation and as a vaccine for diseases such as influenza, graft
CC versus host disease, IgE-mediated allergic reactions, anaphylaxis, adult
CC respiratory distress syndrome (ARDS), Crohn's disease, allergic asthma,
CC acute lymphoblastic leukaemia, non-Hodgkin's lymphoma, Grave's disease,
CC systemic lupus erythematosus, inflammatory immune diseases, myasthenia
CC gravis, immunoproliferative disease lymphadenopathy,
CC angioimmunoproliferative lymphadenopathy, immunoblastive lymphadenopathy,
CC rheumatoid arthritis, diabetes, multiple sclerosis, Alzheimer's disease,
CC osteoporosis and infectious diseases. The antigens are modified to posses
CC a cleavage site (enterokinase or factor Xa) and a Cysteine- containing N-
CC or C-terminal linker peptide which serves as the attachment point to a
CC virus like particle or bacterial protein (the scaffold protein). The
CC present sequence is a cysteine-containing linker peptide used in the
CC molecular antigen array
XX
SQ Sequence 18 AA;

Query Match 100.0%; Score 56; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 1 PKPSTPPGSS 10

RESULT 11
ABR56409
ID ABR56409 standard; peptide; 18 AA.
XX
AC ABR56409;
XX
DT 28-JUL-2003 (first entry)
XX
DE Peptide linker #9.
XX
KW Antigen presenting cell; APC; immune response; virus like particle; VLP;
KW cytostatic; virucide; antibacterial; antiparasitic; fungicide;
KW antiallergic; immunosuppressive; antiaddictive; antiinflammatory;

KW antithyroid; antidiabetic; neuroprotective; nootropic; osteopathic;
KW antirheumatic; antiarthritic; vaccine; immunisation; infectious disease;
KW anti-viral protection; tumour; allergy; drug addiction; Crohn's disease;
KW graft-versus-host disease; Grave's disease; diabetes; multiple sclerosis;
KW Alzheimer's disease; osteoporosis; rheumatoid arthritis;
KW inflammatory autoimmune disease.
XX
OS Synthetic.
XX
PN WO2003024480-A2.
XX
PD 27-MAR-2003.
XX
PF 16-SEP-2002; 2002WO-IB004252.
XX
PR 14-SEP-2001; 2001US-0318967P.
XX
PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
XX
PI Bachmann MF, Storni T, Lechner F;
XX
DR WPI; 2003-363095/34.
XX
PT A composition, useful for enhancing an immune response against an antigen
PT or a virus-like particle, enhancing anti-viral protection in an animal,
PT or immunizing or treating tumors or infectious diseases, e.g. viral
PT infections.
XX
PS Disclosure; Page 65; 243pp; English.
XX
CC The present invention describes a composition (C) for enhancing an immune
CC response against an antigen or a virus-like particle in an animal. (C)
CC comprises a virus-like particle (VLP) bound to at least one antigen, or a
CC VLP capable of being recognised by the immune system of the animal. Also
CC described: (1) enhancing an immune response against an antigen or a VLP
CC in an animal comprising introducing (C) into the animal; (2) vaccines
CC comprising (C) together with a pharmaceutical diluent, carrier or
CC excipient; (3) immunising or treating an animal comprising administering
CC the vaccine to the animal, or priming or boosting a T cell response in
CC the animal by administering the vaccine; and (4) enhancing anti-viral
CC protection in an animal comprising introducing (C) into the animal. (C)
CC has cytostatic, virucide, antibacterial, antiparasitic, fungicide,
CC antiallergic, immunosuppressive, antiaddictive, antiinflammatory,
CC antithyroid, antidiabetic, neuroprotective, nootropic, osteopathic,
CC antirheumatic and antiarthritic activities. (C) or the vaccines can be
CC used for enhancing an immune response against an antigen or a VLP in an
CC animal, enhancing anti-viral protection in an animal, or immunising or
CC treating tumours and infectious diseases such as viral, bacterial,
CC parasitic or fungal infections. The vaccine compositions are also useful
CC for preventing or treating allergies, drug addiction, graft-versus-host
CC disease, Crohn's disease, Grave's disease, diabetes, multiple sclerosis,
CC Alzheimer's disease, osteoporosis, rheumatoid arthritis, or inflammatory
CC autoimmune disease. ACC69838 to ACC69852 and ABR56401 to ABR56509
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 18 AA;

Query Match 100.0%; Score 56; DB 6; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 1 PKPSTPPGSS 10

RESULT 12
ABR44510
ID ABR44510 standard; peptide; 18 AA.
XX
AC ABR44510;
XX
DT 25-JUL-2003 (first entry)
XX
DE Peptide linker #9.
XX
KW Immunostimulatory; virus-like particle; bacteriophage; HBV; LCMV;
KW hepatitis B virus; lymphocytic choriomeningitis virus; vaccine;
KW immunostimulant; cytostatic; antiallergic; virucide; antibacterial;
KW immune response; immunisation; allergy; tumour; breast cancer;
KW neuroblastoma; leukaemia; viral disease; influenza; hepatitis; measles;
KW chicken pox; bacterial infection; tuberculosis; pneumonia; syphilis.
XX
OS Synthetic.
XX
PN WO2003024481-A2.
XX
PD 27-MAR-2003.
XX
PF 16-SEP-2002; 2002WO-IB004132.
XX
PR 14-SEP-2001; 2001US-0318994P.
PR 22-APR-2002; 2002US-0374145P.
XX
PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
PA (MAUR/) MAURER P.
PA (TISS/) TISSOT A.
PA (SCHW/) SCHWARZ K.
PA (MEIJ/) MEIJERINK E.
PA (LIPO/) LIPOWSKY G.
PA (PUMP/) PUMPENS P.
PA (CIEL/) CIELENS I.
PA (RENH/) RENHOFA R.
XX
PI Maurer P, Tissot A, Schwarz K, Meijerink E, Lipowsky G;
PI Pumpens P, Cielens I, Renhofa R, Bachmann MF, Storni T;
XX
DR WPI; 2003-354564/33.
XX
PT New compositions comprising immunostimulatory substances packaged into
PT virus-like particles, useful as a vaccine for enhancing an immune
PT response in animals, e.g. for treating or preventing allergies, tumors or
PT viral infections.
XX
PS Disclosure; Page 75; 322pp; English.
XX
CC The present invention describes a composition (C) for enhancing an immune
CC response in an animal. (C) comprises a virus-like particle (VLP), and an
CC immunostimulatory substance. The immunostimulatory substance is bound to
CC the VLP. Also described: (1) enhancing an immune response in an animal by
CC introducing (C) into the animal; (2) producing (C) for enhancing an
CC immune response in an animal; (3) vaccines comprising (C) together with a
CC pharmaceutical diluent, carrier or excipient; and (4) immunising or
CC treating an animal by: (a) administering the vaccine to the animal; (b)
CC priming a T cell response in the animal by administering the vaccine; or
CC (c) boosting a T cell response in the animal by administering the
CC vaccine. (C) has immunostimulant, cytostatic, antiallergic, virucide and
CC antibacterial activities. (I) can be used in vaccines for enhancing an
CC immune response in an animal, particularly a mammal or human.
CC Specifically, (C) is useful for enhancing a B cell response, a T cell
CC response, or a cytotoxic T-lymphocyte (CTL) response. (C) or a vaccine
CC comprising (C) can also be used for immunising or treating an animal,
CC e.g. humans, sheep, horses, cattle, pigs, dogs, cats, rats, birds,
CC reptiles or fish. (C) is particularly useful in prophylactic or
CC therapeutic vaccines against allergies, tumours (e.g. breast cancers,
CC neuroblastoma, or leukaemia), viral diseases (e.g. influenza, hepatitis,
CC measles or chicken pox), or bacterial infections (e.g. tuberculosis,
CC pneumonia or syphilis). ACC69790 to ACC69815 and ABR44502 to ABR44612
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 18 AA;

Query Match 100.0%; Score 56; DB 6; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 PKPSTPPGSS 10
              ||||||||||
Db          1 PKPSTPPGSS 10


RESULT 13
ADD24210
ID   ADD24210 standard; peptide; 18 AA.
XX
AC   ADD24210;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Linker peptide 4 related to prion disease vaccines.
XX
KW   vaccine composition; virus-like particle; core particle;
KW   first attachment site; antigen; antigenic determinant; prion protein;
KW   PrP; PrP peptide; vaccine; neuroprotective; antiinflammatory;
KW   prion disease; Bovine Spongiform Encephalopathy; BSE;
KW   Creutzfeldt-Jakob Disease; linker peptide.
XX
OS   Synthetic.
XX
PN   WO2003059386-A2.
XX
PD   24-JUL-2003.
XX
PF   17-JAN-2003; 2003WO-EP000460.
XX
PR   18-JAN-2002; 2002US-00050902.
PR   21-JAN-2002; 2002WO-IB000166.
PR   08-JUL-2002; 2002US-0393725P.
PR   18-JUL-2002; 2002US-0396590P.
XX
PA   (CYTO-) CYTOS BIOTECHNOLOGY AG.
XX
PI   Bachmann M,  Maurer P,  Pellicioli E,  Renner WA;
XX
DR   WPI; 2003-598483/56.
XX
PT   A vaccine composition for preventing or treating prion diseases (e.g.
PT   Creutzfeldt-Jakob Disease) comprises a virus-like particle (e.g. RNA-
PT   phage) and at least one prion protein or peptide bound to the virus-like
PT   particle.
XX
PS   Disclosure; Page 59; 246pp; English.
XX
CC   This invention relates to a novel vaccine composition comprising a virus-
CC   like or a core particle with at least one first attachment site and at
CC   least one antigen or antigenic determinant that is a prion protein (PrP)
CC   or its dimer, or a PrP peptide, the antigen or antigenic determinant
CC   being bound to the virus-like or core particle. The vaccine of the
CC   invention may have neuroprotective or antiinflammatory activity. The
CC   composition is useful as a medicament or in manufacturing a medicament
CC   for the treatment or prevention of prion diseases. The prion diseases may
CC   include Bovine Spongiform Encephalopathy (BSE) or Creutzfeldt-Jakob
CC   Disease. The present sequence is that of a linker peptide which is
CC   related to the invention.
XX
SQ   Sequence 18 AA;

  Query Match             100.0%;  Score 56;  DB 7;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              ||||||||||
Db          1 PKPSTPPGSS 10


RESULT 14
AAY91034
ID   AAY91034 standard; peptide; 42 AA.
XX
AC   AAY91034;
XX
DT   06-SEP-2000  (first entry)
XX
DE   Lipid-tagged-scFv protein N-terminus peptide sequence.
XX
KW   Lipid tagged; LT-scFv; lipid modified proteinaceous molecule;
KW   cell therapy; pharmaceutical; cellular membrane.
XX
OS   Unidentified.
XX
PN   WO200023570-A1.
XX
PD   27-APR-2000.
XX
PF   18-OCT-1999;   99WO-NL000644.
XX
PR   16-OCT-1998;   98EP-00203482.
XX
PA   (UBIS-) U-BISYS BV.
XX
PI   Logtenberg T,  De Kruif CA;
XX
DR   WPI; 2000-339673/29.
XX
PT   Altering the protein content of cellular membranes to produce
PT   pharmaceutically active agents.
XX
PS   Disclosure; Page 20; 63pp; English.
XX
CC   The present invention describes a process (I) for modifying the protein
CC   content of cellular membranes using lipid modified proteinaceous
CC   molecules (lmPM). Cells and particles produced via (I) are used as
CC   pharmaceuticals. For example they may be used for a cell therapy
CC   protocol. (I) provides a novel approach to altering the biochemical
CC   properties of cells (especially their ability to target tissues and
CC   organs). It is a very rapid and efficient process and requires only small
CC   amounts of lmPMs which when integrated into the cells are stable in vivo.
CC   (I) does not involve gene transfer (the protein is supplied directly to
CC   the cells) and does not require the cells to be cultured after
CC   integration of the protein. (I) may be applied to a wide range of cell
CC   types not just primary human cells. The present sequence represents the N
CC   -terminus peptide of lipid-tagged-scFv (LT-scFv) proteins expressed in
CC   the vector pLP2, which is used in the exemplification of the present
CC   invention
XX
SQ   Sequence 42 AA;

  Query Match             100.0%;  Score 56;  DB 3;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PKPSTPPGSS 10
              ||||||||||
Db         30 PKPSTPPGSS 39


RESULT 15
AAR39337
ID   AAR39337 standard; protein; 46 AA.
XX
AC   AAR39337;
XX
DT   25-MAR-2003  (revised)
DT   01-FEB-1994  (first entry)
XX
DE   Intercalating GCN4-leucine zipper.
XX
KW   Expression; single-chain Fv; scFv; pLISC-SE; leucine zipper; linker;
KW   hinge; IgG3; GCN4; cassette; restriction site; intercalating peptide;
KW   intercalation.
XX
```

OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 4. .13
FT /label= IgG3-hinge
FT Protein 14. .46
FT /label= GCN4-zipper
XX
PN WO9315210-A1.
XX
PD 05-AUG-1993.
XX
PF 15-JAN-1993; 93WO-EP000082.
XX
PR 23-JAN-1992; 92EP-00101069.
XX
PA (MERE) MERCK PATENT GMBH.
XX
PI Plueckthun A, Pack P;
XX
DR WPI; 1993-258685/32.
DR N-PSDB; AAQ46824.
XX
PT Monomeric and dimeric antibody fragment fusion proteins - that use Fv
PT fragments of antibody but not constant antibody domains.
XX
PS Example 2; Page 30; 44pp; English.
XX
CC Example 2 describes the construction of a gene cassette encoding
CC intercalating peptides of a leucine zipper. The gene cassette, fitted
CC with restriction sites to be compatible with the restriction site at the
CC 3' end of the single-chain (sc) Fv fragment gene, must encode the
CC sequence of a hinge (connecting the scFv fragment to the intercalating
CC peptide) and the intercalating peptide itself. The hinge region, may
CC however be omitted. As an example, the sequence of the upper hinge region
CC of mouse IgG3, followed by the sequence of the leucine zipper sequence of
CC yeast protein GCN4 is back-translated into frequently used E. coli codons
CC (AAQ46824). Oligonucleotides are synthesised and ligated into pLISC-SE
CC (AAQ46823), previously digested with EcoRI and HindIII. (Updated on 25-
CC MAR-2003 to correct PN field.)
XX
SQ Sequence 46 AA;

Query Match 100.0%; Score 56; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSTPPGSS 10
||||||||||
Db 4 PKPSTPPGSS 13

Search completed: April 20, 2004, 10:25:03
Job time : 41.4 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:21:39 ; Search time 44.4 Seconds
(without alignments)
106.594 Million cell updates/sec

Title: US-08-930-480A-5
Perfect score: 84
Sequence: 1 GGGGSGGGGSGGGGS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	84	100.0	170	11	Q925S2	Q925s2 mus musculu
2	84	100.0	241	11	Q921A6	Q921a6 mus musculu
3	84	100.0	243	11	Q7TQM2	Q7tqm2 mus musculu
4	84	100.0	298	11	Q9QYF0	Q9qyf0 mus musculu
5	84	100.0	738	5	O02402	O02402 pinctada fu
6	80	95.2	592	16	Q9PF60	Q9pf60 xylella fas
7	78	92.9	218	11	Q925S1	Q925s1 mus musculu
8	77	91.7	155	5	Q9VZK6	Q9vzk6 drosophila
9	76	90.5	80	10	Q9SUF7	Q9suf7 arabidopsis
10	76	90.5	100	5	Q8MU90	Q8mu90 oncopeltus
11	76	90.5	104	5	Q9GN84	Q9gn84 drosophila
12	76	90.5	104	5	Q9GN83	Q9gn83 drosophila
13	76	90.5	113	10	Q8VY68	Q8vy68 arabidopsis
14	76	90.5	118	5	Q9VYS6	Q9vys6 drosophila
15	76	90.5	158	5	Q9VYD8	Q9vyd8 drosophila
16	76	90.5	175	10	Q9LSN6	Q9lsn6 arabidopsis
17	76	90.5	204	5	Q9W2A7	Q9w2a7 drosophila
18	76	90.5	207	10	Q43522	Q43522 lycopersico
19	76	90.5	221	10	O65514	O65514 arabidopsis
20	76	90.5	222	10	Q7XDV2	Q7xdv2 oryza sativ
21	76	90.5	227	10	Q84W21	Q84w21 arabidopsis
22	76	90.5	255	10	Q9SIH2	Q9sih2 arabidopsis
23	76	90.5	258	16	Q8XLQ8	Q8xlq8 clostridium
24	76	90.5	259	5	Q9V6A4	Q9v6a4 drosophila
25	76	90.5	264	13	Q9DFB6	Q9dfb6 gallus gall
26	76	90.5	272	16	Q7W2S7	Q7w2s7 bordetella
27	76	90.5	280	10	Q7XU62	Q7xu62 oryza sativ
28	76	90.5	283	13	Q8AVB5	Q8avb5 xenopus lae
29	76	90.5	284	4	Q9H524	Q9h524 homo sapien
30	76	90.5	288	10	Q7Y1Z0	Q7y1z0 oryza sativ
31	76	90.5	296	10	Q8RUS0	Q8rus0 arabidopsis
32	76	90.5	300	16	Q7W0G7	Q7w0g7 bordetella
33	76	90.5	304	16	Q7WDS5	Q7wds5 bordetella
34	76	90.5	321	6	Q9MYX6	Q9myx6 ovis aries
35	76	90.5	333	11	Q8C8L2	Q8c8l2 mus musculu
36	76	90.5	333	11	Q7TNS5	Q7tns5 mus musculu
37	76	90.5	335	10	O65330	O65330 elaeagnus u
38	76	90.5	418	5	Q9W2R6	Q9w2r6 drosophila
39	76	90.5	447	13	O73628	O73628 anolis caro
40	76	90.5	493	11	Q8K1I7	Q8k1i7 mus musculu
41	76	90.5	528	4	Q13344	Q13344 homo sapien
42	76	90.5	541	16	Q87BZ7	Q87bz7 xylella fas
43	76	90.5	556	5	Q9VDN4	Q9vdn4 drosophila
44	76	90.5	646	5	Q9VI21	Q9vi21 drosophila
45	76	90.5	647	12	Q65013	Q65013 aleutian mi

ALIGNMENTS

RESULT 1
Q925S2
ID Q925S2 PRELIMINARY; PRT; 170 AA.
AC Q925S2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MRP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX PubMed=11819679;
RA Cui D., Zeng G., Yan X., Wang F., Tian F., Ren D., Zhao T., Li X.,
RA Su C.;
RT "Mechanism of exogenous nucleic acids and their precursors improving
RT the repair of intestinal epithelium after irradiation in mice.";
RL World J. Gastroenterol. 6:709-717(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Cui D., Zeng G., Yan X., Li X., Su C.;
RT "Cloning of mouse genes related to repairing of intestinal epithelium
RT of the irradiated mice by treatment with the intestinal RNA of mice of
RT the same strain.";
RL Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. 19:71-80(2001).
DR EMBL; AF240167; AAK43732.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; ig; 1.
DR SMART; SM00406; IGv; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
SQ SEQUENCE 170 AA; 17978 MW; 5042823CC6C10F38 CRC64;

Query Match 100.0%; Score 84; DB 11; Length 170;
Best Local Similarity 100.0%; Pred. No. 0.03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db        124 GGGGSGGGGSGGGGS 138
```

RESULT 2
Q921A6
ID Q921A6 PRELIMINARY; PRT; 241 AA.
AC Q921A6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Anti-CEA 79 single chain Fv fragment (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98170165; PubMed=9509426;
RA Chung J.H., Choi S.J., Kim H.J., Kim I.J., Choi I.H., Lee S.D.,
RA Yi K.S., Suh P.G., Ryu S.H., Chung H.K.;
RT "Cloning and characterization of cDNAs encoding VH and VL of a
RT monoclonal anti-CEA antibody (CEA 79) cross-reactive with NCA-95 and
RT generation of a single-chain Fv molecule (scFv).";
RL Mol. Cells 7:816-819(1997).
DR EMBL; U88067; AAB48044.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; ig; 2.
DR SMART; SM00406; IGv; 2.
DR PROSITE; PS50835; IG_LIKE; 2.
FT NON_TER 1 1
FT NON_TER 241 241
SQ SEQUENCE 241 AA; 26086 MW; 0276887248E9C771 CRC64;

Query Match 100.0%; Score 84; DB 11; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db        119 GGGGSGGGGSGGGGS 133
```

RESULT 3
Q7TQM2
ID Q7TQM2 PRELIMINARY; PRT; 243 AA.
AC Q7TQM2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ScFv 6H8 protein (Fragment).
GN SCFV 6H8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/C;
RA Peter J.C., Eftekhari P., Billiald P., Wallukat G.;
RT "scFv single chain antibody variable fragment as inverse agonist for
RT the beta-2 adrenergic receptor.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ574851; CAE00495.1; -.
FT NON_TER 1 1
SQ SEQUENCE 243 AA; 25976 MW; BEFFF64D2DCF4F76 CRC64;

Query Match 100.0%; Score 84; DB 11; Length 243;
Best Local Similarity 100.0%; Pred. No. 0.043;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db        117 GGGGSGGGGSGGGGS 131
```

RESULT 4
Q9QYF0
ID Q9QYF0 PRELIMINARY; PRT; 298 AA.
AC Q9QYF0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CN 8 scFv.
GN CN 8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c; TISSUE=Spleen;
RX MEDLINE=20183931; PubMed=10706631;
RA Shinohara N., Demura T., Fukuda H.;
RT "Isolation of a vascular cell wall-specific monoclonal antibody
RT recognizing a cell polarity by using a phage display subtraction
RT method.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:2585-2590(2000).
DR EMBL; AB036341; BAA88633.1; -.
DR PIR; A33933; A33933.
DR PIR; S19112; S19112.
DR HSSP; P01607; 1REI.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; ig; 2.
DR SMART; SM00406; IGv; 2.
DR PROSITE; PS50835; IG_LIKE; 2.
SQ SEQUENCE 298 AA; 31867 MW; E0F96B8A17004317 CRC64;

Query Match 100.0%; Score 84; DB 11; Length 298;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db        158 GGGGSGGGGSGGGGS 172
```

RESULT 5
O02402
ID O02402 PRELIMINARY; PRT; 738 AA.
AC O02402;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Insoluble protein.
OS Pinctada fucata.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Pterioida;
OC Pterioidea; Pteriidae; Pinctada.
OX NCBI_TaxID=50426;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97320490; PubMed=9177341;
RA Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,
RA Nakashima K., Takahashi T.;
RT "Structures of mollusc shell framework proteins.";
RL Nature 387:563-564(1997).
DR EMBL; D86074; BAA20466.1; -.
SQ SEQUENCE 738 AA; 61723 MW; FDF984139BF3BA59 CRC64;

Query Match 100.0%; Score 84; DB 5; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.12;

```
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db        666 GGGGSGGGGSGGGGS 680


RESULT 6
Q9PF60
ID   Q9PF60      PRELIMINARY;      PRT;   592 AA.
AC   Q9PF60;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Endo-1,4-beta-glucanase.
GN   XF0818.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003921; AAF83628.1; -.
DR   PIR; E82759; E82759.
DR   HSSP; P54583; 1ECE.
DR   GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR001547; Glyco_hydro_5.
DR   Pfam; PF00553; CBM_2; 1.
DR   Pfam; PF00150; cellulase; 1.
DR   SMART; SM00637; CBD_II; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW   Complete proteome.
SQ   SEQUENCE   592 AA;  59967 MW;  9846DA4EA3B5C89E CRC64;

  Query Match           95.2%;  Score 80;  DB 16;  Length 592;
  Best Local Similarity  100.0%;  Pred. No. 0.27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              ||||||||||||||
Db        467 GGGGSGGGGSGGGG 480


RESULT 7
Q925S1
ID   Q925S1      PRELIMINARY;      PRT;   218 AA.
AC   Q925S1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MRP5 (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   PubMed=11819679;
RA   Cui D., Zeng G., Yan X., Wang F., Tian F., Ren D., Zhao T., Li X.,
RA   Su C.;
RT   "Mechanism of exogenous nucleic acids and their precursors improving
RT   the repair of intestinal epithelium after irradiation in mice.";
RL   World J. Gastroenterol. 6:709-717(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Cui D., Zeng G., Yan X., Li X., Su C.;
RT   "Cloning of mouse genes related to repairing of intestinal epithelium
RT   of the irradiated mice by treatment with the intestinal RNA of mice of
RT   the same strain.";
RL   Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. 19:71-80(2001).
DR   EMBL; AF240168; AAK43733.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER     218    218
SQ   SEQUENCE   218 AA;  23013 MW;  527E4FA8F7982817 CRC64;

  Query Match           92.9%;  Score 78;  DB 11;  Length 218;
  Best Local Similarity  93.3%;  Pred. No. 0.17;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||.|||
Db        121 GGGGSGGGGSGSGGS 135


RESULT 8
Q9VZK6
ID   Q9VZK6      PRELIMINARY;      PRT;   155 AA.
AC   Q9VZK6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CG10853 protein (LP09837p).
GN   CG10853 OR BCDNA:LP09837.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
```

```
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003479; AAF47815.1; -.
DR   EMBL; AY075437; AAL68252.1; -.
DR   FlyBase; FBgn0035478; CG10853.
SQ   SEQUENCE   155 AA;  14855 MW;  EF7D78EDD16675BF CRC64;

  Query Match             91.7%;  Score 77;  DB 5;  Length 155;
  Best Local Similarity   86.7%;  Pred. No. 0.15;
  Matches   13;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              ||||||||| ||||:
Db         58 GGGGSGGGGGGGGGA 72


RESULT 9
Q9SUF7
ID   Q9SUF7      PRELIMINARY;      PRT;    80 AA.
AC   Q9SUF7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   T12G13.70 OR AT4G08230.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,
RA   Barrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,
RA   Schueller C.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA   Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL080252; CAB45793.1; -.
DR   EMBL; AL161510; CAB81159.1; -.
DR   PIR; T10550; T10550.
KW   Hypothetical protein.
SQ   SEQUENCE   80 AA;  7872 MW;  A1BEEE43FCA7ED68 CRC64;

  Query Match             90.5%;  Score 76;  DB 10;  Length 80;
  Best Local Similarity   92.9%;  Pred. No. 0.1;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         22 GGGGGGGGGSGGGG 35


RESULT 10
Q8MU90
ID   Q8MU90      PRELIMINARY;      PRT;   100 AA.
AC   Q8MU90;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Tiptop transcription factor (Fragment).
GN   TIPTOP.
OS   Oncopeltus fasciatus (Milkweed bug).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Paraneoptera; Hemiptera; Euhemiptera; Heteroptera;
OC   Panheteroptera; Pentatomomorpha; Lygaeoidea; Lygaeidae; Lygaeinae;
OC   Oncopeltus.
OX   NCBI_TaxID=7536;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rogers B.T., Herke S.W.;
RT   "Partial tiptop cDNA from Oncopeltus fasciatus.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF533539; AAM97356.1; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
FT   NON_TER       1      1
FT   NON_TER     100    100
SQ   SEQUENCE   100 AA;  9919 MW;  650D5E401FEF35CD CRC64;

  Query Match             90.5%;  Score 76;  DB 5;  Length 100;
  Best Local Similarity   86.7%;  Pred. No. 0.13;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db         65 GGGGGGGGGGGGGGS 79


RESULT 11
```

```
Q9GN84
ID   Q9GN84      PRELIMINARY;      PRT;   104 AA.
AC   Q9GN84;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   NONA protein (Fragment).
GN   NONA.
OS   Drosophila littoralis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=47316;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ru3, and Ru1;
RA   Huttunen S., Campesan S., Hoikkala A.;
RT   "Intra- and interspecific nucleotide variation at the nonA gene in
RT   Drosophila littoralis and D. virilis.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ304367; CAC20090.1; -.
DR   EMBL; AJ304361; CAC20084.1; -.
DR   FlyBase; FBgn0043410; Dlit\nonA.
FT   NON_TER     104    104
SQ   SEQUENCE   104 AA;  10048 MW;  AC804E039196298C CRC64;

  Query Match             90.5%;  Score 76;  DB 5;  Length 104;
  Best Local Similarity   92.9%;  Pred. No. 0.13;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         68 GGGGGGGGGSGGGG 81


RESULT 12
Q9GN83
ID   Q9GN83      PRELIMINARY;      PRT;   104 AA.
AC   Q9GN83;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   NONA protein (Fragment).
GN   NONA.
OS   Drosophila littoralis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=47316;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sa5, Ou5, Ku5, Sa1, and Sa3;
RA   Huttunen S., Campesan S., Hoikkala A.;
RT   "Intra- and interspecific nucleotide variation at the nonA gene in
RT   Drosophila littoralis and D. virilis.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ304352; CAC20075.1; -.
DR   EMBL; AJ304316; CAC20039.1; -.
DR   EMBL; AJ304334; CAC20057.1; -.
DR   EMBL; AJ304340; CAC20063.1; -.
DR   EMBL; AJ304346; CAC20069.1; -.
DR   FlyBase; FBgn0043410; Dlit\nonA.
FT   NON_TER     104    104
SQ   SEQUENCE   104 AA;  10047 MW;  AC8EAE039196298C CRC64;

  Query Match             90.5%;  Score 76;  DB 5;  Length 104;
  Best Local Similarity   92.9%;  Pred. No. 0.13;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         68 GGGGGGGGGSGGGG 81


RESULT 13
Q8VY68
ID   Q8VY68      PRELIMINARY;      PRT;   113 AA.
AC   Q8VY68;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   AT4G08230.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY072406; AAL62398.1; -.
DR   EMBL; BT000222; AAN15541.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   113 AA;  11548 MW;  D0182159545EF3F9 CRC64;

  Query Match             90.5%;  Score 76;  DB 10;  Length 113;
  Best Local Similarity   92.9%;  Pred. No. 0.14;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         63 GGGGGGGGGSGGGG 76


RESULT 14
Q9VYS6
ID   Q9VYS6      PRELIMINARY;      PRT;   118 AA.
AC   Q9VYS6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CG1840 protein (LD12750p).
GN   CG1840 OR BCDNA:LD12750.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
```

```
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003487; AAF48112.1; -.
DR   EMBL; AY094783; AAM11136.1; -.
DR   FlyBase; FBgn0030351; CG1840.
SQ   SEQUENCE   118 AA;  11803 MW;  0568ACA6501716AC CRC64;

  Query Match             90.5%;  Score 76;  DB 5;  Length 118;
  Best Local Similarity   86.7%;  Pred. No. 0.15;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db         74 GGGGGGGGGGGGGGS 88


RESULT 15
Q9VYD8
ID   Q9VYD8      PRELIMINARY;      PRT;   158 AA.
AC   Q9VYD8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG1987 protein (RE47308p).
GN   RBP1-LIKE OR CG1987.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
```

RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE003492; AAF48264.2; -.
DR EMBL; AY113490; AAM29495.1; -.
DR FlyBase; FBgn0030479; Rbp1-like.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
SQ SEQUENCE 158 AA; 16799 MW; C6D065ABD640EAE5 CRC64;

Query Match 90.5%; Score 76; DB 5; Length 158;
Best Local Similarity 92.9%; Pred. No. 0.2;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         94 GGGGGGGGGSGGGG 107
```

Search completed: April 20, 2004, 10:27:02
Job time : 46.4 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:21:04 ; Search time 12.6 Seconds
(without alignments)
61.988 Million cell updates/sec

Title: US-08-930-480A-5
Perfect score: 84
Sequence: 1 GGGGSGGGGSGGGGS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	80	95.2	306	1	RALY_HUMAN	Q9ukm9 homo sapien
2	79	94.0	104	1	HOL3_HOLDI	Q25055 holotrichia
3	77	91.7	206	1	TWS1_MOUSE	P26687 mus musculu
4	76	90.5	165	1	GRP1_ORYSA	P25074 oryza sativ
5	76	90.5	280	1	CHIA_MAIZE	P29022 zea mays (m
6	76	90.5	321	1	PUR_MOUSE	P42669 mus musculu
7	76	90.5	322	1	PUR_HUMAN	Q00577 homo sapien
8	76	90.5	378	1	RU17_MOUSE	Q62376 mus musculu
9	76	90.5	433	1	HXB3_MOUSE	P09026 mus musculu
10	76	90.5	539	1	DOP2_DROME	Q24563 drosophila
11	76	90.5	588	1	T7L1_HUMAN	Q9hcs4 homo sapien
12	76	90.5	622	1	SR68_CANFA	Q00004 canis famil
13	76	90.5	663	1	DUS8_MOUSE	O09112 mus musculu
14	76	90.5	688	1	EOMD_MOUSE	O54839 mus musculu
15	76	90.5	979	1	RFX1_HUMAN	P22670 homo sapien
16	76	90.5	1627	1	TP2B_CHICK	O42131 gallus gall
17	76	90.5	3703	1	ABF1_HUMAN	Q15911 homo sapien
18	76	90.5	3726	1	ABF1_MOUSE	Q61329 mus musculu
19	74	88.1	266	1	CANS_RABIT	P06813 oryctolagus
20	74	88.1	331	1	SHX2_MOUSE	P70390 mus musculu
21	74	88.1	440	1	FXGA_CHICK	Q98937 gallus gall
22	74	88.1	465	1	FXD3_MOUSE	Q61060 mus musculu
23	74	88.1	627	1	SR68_HUMAN	Q9uhb9 homo sapien
24	74	88.1	722	1	COAT_PAVHH	P03136 hamster par
25	74	88.1	895	1	Z281_HUMAN	Q9y2x9 homo sapien
26	73	86.9	151	1	GDF7_MOUSE	P43029 mus musculu
27	73	86.9	263	1	CANS_BOVIN	P13135 bos taurus
28	73	86.9	266	1	CANS_PIG	P04574 sus scrofa
29	73	86.9	268	1	CANS_HUMAN	P04632 homo sapien
30	73	86.9	367	1	BET3_MESAU	O09029 mesocricetu
31	73	86.9	377	1	HSF7_ARATH	Q9t0d3 arabidopsis
32	73	86.9	387	1	SOX1_HUMAN	O00570 homo sapien
33	73	86.9	391	1	SOX1_MOUSE	P53783 mus musculu
34	73	86.9	403	1	MAF_HUMAN	O75444 homo sapien
35	73	86.9	476	1	EVX2_HUMAN	Q03828 homo sapien
36	73	86.9	495	1	PO33_MOUSE	P31361 mus musculu
37	73	86.9	497	1	PO33_RAT	Q63262 rattus norv
38	73	86.9	500	1	PO33_HUMAN	P20264 homo sapien
39	73	86.9	933	1	NPA3_HUMAN	Q8ixf0 homo sapien
40	72	85.7	141	1	YORE_TTV1	P19289 thermoprote
41	72	85.7	199	1	AC22_TENMO	P26968 tenebrio mo
42	72	85.7	282	1	TR2A_HUMAN	Q13595 homo sapien
43	72	85.7	303	1	CHIB_POPTR	P29031 populus tri
44	72	85.7	331	1	SHX2_HUMAN	O60902 homo sapien
45	72	85.7	348	1	SXL_CERCA	O61374 ceratitis c

ALIGNMENTS

```
RESULT 1
RALY_HUMAN
ID   RALY_HUMAN      STANDARD;      PRT;   306 AA.
AC   Q9UKM9; Q14621; Q9BQX6; Q9UJB3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   RNA-binding protein Raly (hnRNP associated with lethal yellow homolog)
DE   (Autoantigen p542).
GN   RALY OR P452.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Testis;
RX   MEDLINE=99431566; PubMed=10500250;
RA   Khrebtukova I., Kuklin A., Woychik R.P., Michaud E.J.;
RT   "Alternative processing of the human and mouse raly genes.";
RL   Biochim. Biophys. Acta 1447:107-112(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Vaughan J.H.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21638749; PubMed=11780052;
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [4]
```

```
RP   SEQUENCE OF 85-306 FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.
RC   TISSUE=Lymphocytes;
RX   MEDLINE=98018738; PubMed=9376072;
RA   Rhodes G.H., Valbracht J.R., Nguyen M.-D., Vaughan J.H.;
RT   "The p542 gene encodes an autoantigen that cross-reacts with EBNA-1 of
RT   the Epstein Barr virus and which may be a heterogeneous nuclear
RT   ribonucleoprotein.";
RL   J. Autoimmun. 10:447-454(1997).
RN   [5]
RP   SEQUENCE OF 227-253, AND DETERMINATION OF AUTOANTIGENIC EPITOPE.
RX   MEDLINE=95190029; PubMed=7533788;
RA   Vaughan J.H., Valbracht J.R., Nguyen M.-D., Handley H.H., Smith R.S.,
RA   Patrick K., Rhodes G.H.;
RT   "Epstein-Barr virus-induced autoimmune responses. I. Immunoglobulin M
RT   autoantibodies to proteins mimicking and not mimicking Epstein-Barr
RT   virus nuclear antigen-1.";
RL   J. Clin. Invest. 95:1306-1315(1995).
CC   -!- FUNCTION: Probable-RNA binding protein. Could be a heterogeneous
CC       nuclear ribonucleoprotein (hnRNP).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=2;
CC         IsoId=Q9UKM9-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=Q9UKM9-2; Sequence=VSP_005804;
CC   -!- TISSUE SPECIFICITY: Expressed in heart, brain, lung, liver,
CC       skeletal muscle, kidney and pancreas. Weakly expressed in
CC       placenta.
CC   -!- DISEASE: Autoantigen found in infectious mononucleosis caused by
CC       Epstein-Barr virus. An epitope recognized by B-cells, which cross-
CC       react with the BKRF1 protein (EBNA-1 nuclear protein) of Epstein-
CC       Barr virus has been identified.
CC   -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC   -!- CAUTION: Ref.4 (CAC29371) sequence differs from that shown due to
CC       erroneous gene model prediction.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF148457; AAF04487.1; -.
DR   EMBL; L38696; AAC28898.1; -.
DR   EMBL; AL031668; CAC29371.1; ALT_SEQ.
DR   EMBL; AL031668; CAB43742.1; -.
DR   Genew; HGNC:15921; RALY.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Ribonucleoprotein; RNA-binding; Nuclear protein; Antigen;
KW   Alternative splicing; Polymorphism.
FT   DOMAIN       21     92       RNA-BINDING (RRM).
FT   DOMAIN      227    253       EPITOPE (RECOGNIZED BY BKRF1 ANTIBODIES).
FT   DOMAIN      225    251       POLY-GLY.
FT   VARSPLIC    110    125       Missing (in isoform 1).
FT                                /FTId=VSP_005804.
FT   VARIANT     215    215       Q -> R (in dbSNP:3180568).
FT                                /FTId=VAR_015223.
FT   VARIANT     251    251       G -> S (in dbSNP:2281209).
FT                                /FTId=VAR_015224.
FT   CONFLICT    214    215       EQ -> DE (IN REF. 2).
FT   CONFLICT    230    230       A -> AS (IN REF. 2).
SQ   SEQUENCE   306 AA;  32463 MW;  7F4376D3BD8E4728 CRC64;

  Query Match             95.2%;  Score 80;  DB 1;  Length 306;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              ||||||||||||||
Db        235 GGGGSGGGGSGGGG 248


RESULT 2
HOL3_HOLDI
ID   HOL3_HOLDI     STANDARD;      PRT;   104 AA.
AC   Q25055;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Holotricin 3 precursor.
OS   Holotrichia diomphalia.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Coleoptera; Polyphaga; Scarabaeiformia;
OC   Scarabaeidae; Melolonthinae; Holotrichia.
OX   NCBI_TaxID=33394;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 21-40.
RC   TISSUE=Larval hemolymph;
RX   MEDLINE=96073722; PubMed=8535393;
RA   Lee S.Y., Moon H.J., Kurata S., Natori S., Lee B.L.;
RT   "Purification and cDNA cloning of an antifungal protein from the
RT   hemolymph of Holotrichia diomphalia larvae.";
RL   Biol. Pharm. Bull. 18:1049-1052(1995).
CC   -!- FUNCTION: Has antifungal activity against C.albicans.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: TO TENECIN 3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13744; BAA02889.1; -.
DR   PIR; JC4190; JC4190.
KW   Insect immunity; Antibiotic; Hemolymph; Fungicide; Signal; Repeat.
FT   SIGNAL        1     20
FT   CHAIN        21    104       HOLOTRICIN 3.
FT   DOMAIN       27     98       18 X 4 AA APPROXIMATE TANDEM REPEATS OF
FT                                H-G-G-G.
FT   REPEAT       27     30       1.
FT   REPEAT       31     34       2.
FT   REPEAT       35     38       3.
FT   REPEAT       39     42       4.
FT   REPEAT       43     46       5.
FT   REPEAT       47     50       6.
FT   REPEAT       51     54       7.
FT   REPEAT       55     58       8.
FT   REPEAT       59     62       9.
FT   REPEAT       63     66       10.
FT   REPEAT       67     70       11.
FT   REPEAT       71     74       12.
FT   REPEAT       75     78       13.
FT   REPEAT       79     82       14.
FT   REPEAT       83     86       15.
FT   REPEAT       87     90       16.
FT   REPEAT       91     94       17.
FT   REPEAT       96     98       18.
SQ   SEQUENCE   104 AA;  9026 MW;  2799D681BFDCC725 CRC64;

  Query Match             94.0%;  Score 79;  DB 1;  Length 104;
  Best Local Similarity   93.3%;  Pred. No. 0.061;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| ||||||||||
Db         64 GGGGRGGGGSGGGGS 78
```

```
RESULT 3
TWS1_MOUSE
ID   TWS1_MOUSE     STANDARD;      PRT;   206 AA.
AC   P26687;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Twist related protein 1 (M-twist).
GN   TWIST1 OR TWIST.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91122450; PubMed=1840517;
RA   Wolf C., Thisse C., Stoetzel C., Thisse B., Gerlinger P.,
RA   Perrin-Schmitt F.;
RT   "The M-twist gene of Mus is expressed in subsets of mesodermal cells
RT   and is closely related to the Xenopus X-twi and the Drosophila twist
RT   genes.";
RL   Dev. Biol. 143:363-373(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=98001585; PubMed=9343420;
RA   Hamamori Y., Wu H.Y., Sartorelli V., Kedes L.;
RT   "The basic domain of myogenic basic helix-loop-helix (bHLH) proteins
RT   is the novel target for direct inhibition by another bHLH protein,
RT   Twist.";
RL   Mol. Cell. Biol. 17:6563-6573(1997).
RN   [4]
RP   FUNCTION, AND INDUCTION.
RX   MEDLINE=22441800; PubMed=12553906;
RA   Sosic D., Richardson J.A., Yu K., Ornitz D.M., Olson E.N.;
RT   "Twist regulates cytokine gene expression through a negative feedback
RT   loop that represses NF-kappaB activity.";
RL   Cell 112:169-180(2003).
CC   -!- FUNCTION: Probable transcription factor, which seems to be
CC       involved in the negative regulation of cellular determination and
CC       in the differentiation of several lineages including myogenesis,
CC       osteogenesis, and neurogenesis. Inhibits myogenesis by
CC       sequestrating E proteins, inhibiting trans-activation by Mef2, and
CC       inhibiting DNA-binding by Myod through physical interaction. This
CC       interaction probably involves the basic domains of both proteins.
CC       Also represses expression of proinflammatory cytokines such as
CC       Tnfa and Il1b.
CC   -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC       bHLH protein. Homodimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Subset of mesodermal cells.
CC   -!- INDUCTION: By TNF-alpha.
CC   -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63649; AAA40514.1; -.
DR   EMBL; M63650; AAA40515.1; -.
DR   EMBL; BC033434; AAH33434.1; -.
DR   PIR; I53066; I53066.
DR   TRANSFAC; T01635; -.
DR   MGD; MGI:98872; Twist1.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0030154; P:cell differentiation; IMP.
DR   GO; GO:0030326; P:limb morphogenesis; IMP.
DR   GO; GO:0045596; P:negative regulation of cell differentiation; IDA.
DR   GO; GO:0045843; P:negative regulation of myogenesis; IDA.
DR   GO; GO:0001679; P:neurulation; IMP.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001092; HLH_basic.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   PROSITE; PS50888; HLH; 1.
KW   Differentiation; Developmental protein; Nuclear protein; DNA-binding;
KW   Transcription regulation.
FT   DOMAIN       80    102       GLY-RICH.
FT   DNA_BIND     112   124       BASIC DOMAIN.
FT   DOMAIN       125   164       HELIX-LOOP-HELIX MOTIF.
FT   VARIANT       36    36       A -> R (IN CDNA).
FT   VARIANT       91    91       G -> P (IN CDNA).
SQ   SEQUENCE   206 AA;  21198 MW;  618AD8E9BE87C555 CRC64;

  Query Match             91.7%;  Score 77;  DB 1;  Length 206;
  Best Local Similarity   86.7%;  Pred. No. 0.17;
  Matches   13;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              ||||:|||| |||||
Db         83 GGGGAGGGGGGGGGS 97


RESULT 4
GRP1_ORYSA
ID   GRP1_ORYSA     STANDARD;      PRT;   165 AA.
AC   P25074;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Glycine-rich cell wall structural protein 1 precursor.
GN   GRP-1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Indica-IR36;
RX   MEDLINE=91370862; PubMed=1716496;
RA   Lei M., Wu R.;
RT   "A novel glycine-rich cell wall protein gene in rice.";
RL   Plant Mol. Biol. 16:187-198(1991).
CC   -!- FUNCTION: Responsible for plasticity of the cell wall (Potential).
CC   -!- SUBCELLULAR LOCATION: Cell wall (Potential).
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X53596; CAA37665.1; -.
DR   PIR; S13385; KNRZG1.
DR   Gramene; P25074; -.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    165       GLYCINE-RICH CELL WALL STRUCTURAL
FT                                 PROTEIN 1.
FT   DOMAIN       31    159       GLY-RICH.
FT   REPEAT       56     62       R2 (TYR-RICH).
FT   REPEAT       93     99       R2 (TYR-RICH).
FT   REPEAT      132    138       R2 (TYR-RICH).
SQ   SEQUENCE   165 AA;  13536 MW;  E36CE31C3650AC9A CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 165;
  Best Local Similarity   86.7%;  Pred. No. 0.18;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db        111 GGGGGGGGGGGGGGS 125


RESULT 5
CHIA_MAIZE
ID   CHIA_MAIZE     STANDARD;      PRT;   280 AA.
AC   P29022;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Endochitinase A precursor (EC 3.2.1.14) (Seed chitinase A).
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seed;
RX   MEDLINE=92202208; PubMed=1551872;
RA   Huynh Q.K., Hironaka C.M., Levine E.B., Smith C.E., Borgmeyer J.R.,
RA   Shah D.M.;
RT   "Antifungal proteins from plants. Purification, molecular cloning,
RT   and antifungal properties of chitinases from maize seed.";
RL   J. Biol. Chem. 267:6635-6640(1992).
RN   [2]
RP   SEQUENCE OF 180-195.
RC   TISSUE=Seed;
RX   MEDLINE=92156129; PubMed=1740436;
RA   Verburg J.G., Smith C.E., Lisek C.A., Huynh Q.K.;
RT   "Identification of an essential tyrosine residue in the catalytic
RT   site of a chitinase isolated from Zea mays that is selectively
RT   modified during inactivation with
RT   1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide.";
RL   J. Biol. Chem. 267:3886-3893(1992).
CC   -!- FUNCTION: This protein functions as a defense against chitin
CC       containing fungal pathogens.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC       acetyl-D-glucosamine polymers of chitin.
CC   -!- MISCELLANEOUS: Maize chitinase B seems to be less active than
CC       chitinase A.
CC   -!- SIMILARITY: Belongs to chitinase class IA (family 19 of glycosyl
CC       hydrolases).
CC   -!- SIMILARITY: Contains 1 chitin-binding type-1 domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M84164; AAA33444.1; -.
DR   HSSP; P02877; 1HEV.
DR   MaizeDB; 25130; -.
DR   InterPro; IPR001002; Chitin_binding_1.
DR   InterPro; IPR000726; Glyco_hydro_19.
DR   Pfam; PF00187; Chitin_bind_1; 1.
DR   Pfam; PF00182; Glyco_hydro_19; 1.
DR   PRINTS; PR00451; CHITINBINDNG.
DR   ProDom; PD000609; Chitin_binding_1; 1.
DR   ProDom; PD354900; Glyco_hydro_19; 1.
DR   SMART; SM00270; ChtBD1; 1.
DR   PROSITE; PS50941; CHIT_BIND_I_2; 1.
DR   PROSITE; PS00026; CHIT_BIND_I_1; 1.
DR   PROSITE; PS00773; CHITINASE_19_1; 1.
DR   PROSITE; PS00774; CHITINASE_19_2; 1.
KW   Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal;
KW   Multigene family.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    280       ENDOCHITINASE A.
FT   DOMAIN       26     60       CHITIN-BINDING TYPE-1.
FT   DOMAIN       61     77       HINGE REGION (POLY-GLY).
FT   DOMAIN       78    280       CATALYTIC.
FT   DISULFID     28     36       BY SIMILARITY.
FT   DISULFID     30     42       BY SIMILARITY.
FT   DISULFID     35     49       BY SIMILARITY.
FT   DISULFID     53     58       BY SIMILARITY.
FT   DISULFID    100    149       BY SIMILARITY.
FT   DISULFID    161    170       BY SIMILARITY.
FT   DISULFID    248    280       BY SIMILARITY.
SQ   SEQUENCE   280 AA;  29124 MW;  4FC5BB7D938C1CC1 CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 280;
  Best Local Similarity   86.7%;  Pred. No. 0.28;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db         61 GGGGGGGGGGGGGGS 75


RESULT 6
PUR_MOUSE
ID   PUR_MOUSE      STANDARD;      PRT;   321 AA.
AC   P42669;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Transcriptional activator protein PUR-alpha (Purine-rich single-
DE   stranded DNA-binding protein alpha).
GN   PURA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95047494; PubMed=7959008;
RA   Ma Z.-W., Bergemann A.D., Johnson E.M.;
RT   "Conservation in human and mouse Pur alpha of a motif common to
RT   several proteins involved in initiation of DNA replication.";
RL   Gene 149:311-314(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX   MEDLINE=97476282; PubMed=9334258;
```

```
RA   Kelm R.J. Jr., Elder P.K., Strauch A.R., Getz M.J.;
RT   "Sequence of cDNAs encoding components of vascular actin
RT   single-stranded DNA-binding factor 2 establish identity to Puralpha
RT   and Purbeta.";
RL   J. Biol. Chem. 272:26727-26733(1997).
CC   -!- FUNCTION: THIS IS A PROBABLE TRANSCRIPTION ACTIVATOR THAT
CC       SPECIFICALLY BINDS THE PURINE-RICH SINGLE STRAND OF THE PUR
CC       ELEMENT LOCATED UPSTREAM OF THE C-MYC GENE. MAY PLAY A ROLE IN
CC       THE INITIATION OF DNA REPLICATION AND IN RECOMBINATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U02098; AAA64630.1; -.
DR   EMBL; AF017631; AAB71860.1; -.
DR   TRANSFAC; T05167; -.
DR   MGD; MGI:103079; Pura.
DR   InterPro; IPR006628; PUR_DNA_RNA.
DR   Pfam; PF04845; PurA; 1.
DR   SMART; SM00712; PUR; 3.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein.
FT   DOMAIN       11     52       GLY-RICH.
FT   DOMAIN      292    321       GLN/GLU-RICH; PART OF THE TRANSCRIPTIONAL
FT                                ACTIVATION DOMAIN (POTENTIAL).
SQ   SEQUENCE   321 AA;  34884 MW;  0379DBD96D47DCEA CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 321;
  Best Local Similarity   92.9%;  Pred. No. 0.32;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         34 GGGGGGGGGSGGGG 47


RESULT 7
PUR_HUMAN
ID   PUR_HUMAN      STANDARD;      PRT;   322 AA.
AC   Q00577;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional activator protein PUR-alpha (Purine-rich single-
DE   stranded DNA-binding protein alpha).
GN   PURA OR PUR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal liver;
RX   MEDLINE=93078769; PubMed=1448097;
RA   Bergemann A.D., Ma Z.-W., Johnson E.M.;
RT   "Sequence of cDNA comprising the human pur gene and sequence-specific
RT   single-stranded-DNA-binding properties of the encoded protein.";
RL   Mol. Cell. Biol. 12:5673-5682(1992).
CC   -!- FUNCTION: This is a probable transcription activator that
CC       specifically binds the purine-rich single strand of the PUR
CC       element located upstream of the MYC gene. May play a role in
CC       the initiation of DNA replication and in recombination.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M96684; AAA60229.1; -.
DR   EMBL; U02098; -; NOT_ANNOTATED_CDS.
DR   Genew; HGNC:9701; PURA.
DR   MIM; 600473; -.
DR   GO; GO:0003705; F:RNA polymerase II transcription factor acti. . .; TAS.
DR   GO; GO:0003697; F:single-stranded DNA binding; TAS.
DR   GO; GO:0006270; P:DNA replication initiation; TAS.
DR   InterPro; IPR006628; PUR_DNA_RNA.
DR   Pfam; PF04845; PurA; 1.
DR   SMART; SM00712; PUR; 3.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein.
FT   DOMAIN       11     53       GLY-RICH.
FT   DOMAIN      293    322       GLN/GLU-RICH; PART OF THE TRANSCRIPTIONAL
FT                                ACTIVATION DOMAIN (POTENTIAL).
SQ   SEQUENCE   322 AA;  34911 MW;  797968504D01B356 CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 322;
  Best Local Similarity   92.9%;  Pred. No. 0.32;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         34 GGGGGGGGGSGGGG 47


RESULT 8
RU17_MOUSE
ID   RU17_MOUSE     STANDARD;      PRT;   378 AA.
AC   Q62376;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   U1 small nuclear ribonucleoprotein 70 kDa (U1 SNRNP 70 kDa) (snRNP70)
DE   (Fragment).
GN   SNRP70.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   STRAIN=BALB/c;
RX   MEDLINE=89276388; PubMed=2525092;
RA   Hornig H., Fischer U., Costas M., Rauh A., Luehrmann R.;
RT   "Analysis of genomic clones of the murine U1RNA-associated 70-kDa
RT   protein reveals a high evolutionary conservation of the protein
RT   between human and mouse.";
RL   Eur. J. Biochem. 182:45-50(1989).
CC   -!- FUNCTION: Mediates the splicing of pre-mRNA by binding to the loop
CC       I region of U1-snRNA. The truncated isoform cannot bind U1-snRNA
CC       (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q62376-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q62376-2; Sequence=VSP_005851, VSP_005852;
CC   -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE C-
CC       TERMINAL REGION (BY SIMILARITY).
CC   -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; X15769; CAA33777.1; -.
DR   EMBL; X15770; CAA33777.1; JOINED.
DR   EMBL; X15771; CAA33777.1; JOINED.
DR   EMBL; X15772; CAA33777.1; JOINED.
DR   EMBL; X15774; CAA33777.1; JOINED.
DR   EMBL; X15775; CAA33777.1; JOINED.
DR   EMBL; X15776; CAA33777.1; JOINED.
DR   PIR; S04336; S04336.
DR   PIR; S04824; S04824.
DR   HSSP; P09651; 1HA1.
DR   MGD; MGI:98341; Snrp70.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Nuclear protein; Ribonucleoprotein; RNA-binding; Phosphorylation;
KW   Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN       33    111       RNA-BINDING (RRM).
FT   DOMAIN      161    240       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      241    256       POLY-GLY.
FT   DOMAIN      286    333       ARG/ASP/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      334    339       POLY-GLY.
FT   VARSPLIC     90     96       AYKHADG -> TTQLACS (in isoform 2).
FT                                /FTId=VSP_005851.
FT   VARSPLIC     97    378       Missing (in isoform 2).
FT                                /FTId=VSP_005852.
SQ   SEQUENCE   378 AA;  43722 MW;  E669C31BCA365AA0 CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 378;
  Best Local Similarity   92.9%;  Pred. No. 0.36;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db        241 GGGGGGGGGSGGGG 254


RESULT 9
HXB3_MOUSE
ID   HXB3_MOUSE     STANDARD;      PRT;   433 AA.
AC   P09026; P10285; Q61680;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein Hox-B3 (Hox-2.7) (MH-23).
GN   HOXB3 OR HOXB-3 OR HOX-2.7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92258392; PubMed=1582411;
RA   Sham M.H., Hunt P., Nonchev S., Papalopulu N., Graham A.,
RA   Boncinelli E., Krumlauf R.;
RT   "Analysis of the murine Hox-2.7 gene: conserved alternative
RT   transcripts with differential distributions in the nervous system and
RT   the potential for shared regulatory regions.";
RL   EMBO J. 11:1825-1836(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95196953; PubMed=7890121;
RA   Brown W.M., Taylor G.R.;
RT   "The 5'-sequence of the murine Hox-b3 (Hox-2.7) gene and its intron
RT   contain multiple transcription-regulatory elements.";
RL   Int. J. Biochem. 26:1403-1409(1994).
RN   [3]
RP   SEQUENCE OF 152-361 FROM N.A.
RX   MEDLINE=88054465; PubMed=2890503;
RA   Lonai P., Arman E., Czosnek H., Ruddle F.H., Blatt C.;
RT   "New murine homeoboxes: structure, chromosomal assignment, and
RT   differential expression in adult erythropoiesis.";
RL   DNA 6:409-418(1987).
RN   [4]
RP   SEQUENCE OF 181-265 FROM N.A.
RX   MEDLINE=89091992; PubMed=2463210;
RA   Graham A., Papalopulu N., Lorimer J., McVey J.H., Tuddenham E.G.D.,
RA   Krumlauf R.;
RT   "Characterization of a murine homeo box gene, Hox-2.6, related to the
RT   Drosophila Deformed gene.";
RL   Genes Dev. 2:1424-1438(1988).
CC   -!- FUNCTION: Sequence-specific transcription factor which is part of
CC       a developmental regulatory system that provides cells with
CC       specific positional identities on the anterior-posterior axis.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the Antp homeobox family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X66177; CAA46951.1; -.
DR   EMBL; U02278; AAB60496.1; -.
DR   EMBL; M18168; AAA37840.1; -.
DR   PIR; S20963; S20963.
DR   HSSP; P02833; 1SAN.
DR   TRANSFAC; T01724; -.
DR   MGD; MGI:96184; Hoxb3.
DR   InterPro; IPR001827; Antennapedia.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00025; ANTENNAPEDIA.
DR   PRINTS; PR00024; HOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS00032; ANTENNAPEDIA; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW   Transcription regulation.
FT   SITE        129    134       ANTP-TYPE HEXAPEPTIDE.
FT   DOMAIN      154    181       GLY-RICH.
FT   DNA_BIND    191    250       HOMEOBOX.
FT   CONFLICT    113    113       G -> C (IN REF. 1).
FT   CONFLICT    119    119       A -> S (IN REF. 1).
FT   CONFLICT    152    168       GCGGGGGGGGGGGGGGG -> RLWWWRPAVVAAAAAVR
FT                                (IN REF. 3).
FT   CONFLICT    182    182       D -> N (IN REF. 4).
FT   CONFLICT    216    217       LC -> FV (IN REF. 3).
FT   CONFLICT    330    330       S -> L (IN REF. 3).
FT   CONFLICT    342    361       GAYGTPTMQGSPVYVGGGGY -> APTGRPPCRAVRCMWAG
FT                                VAT (IN REF. 3).
SQ   SEQUENCE   433 AA;  44353 MW;  9AD3C922663612A6 CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 433;
  Best Local Similarity   86.7%;  Pred. No. 0.41;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db        156 GGGGGGGGGGGGGGS 170


RESULT 10
DOP2_DROME
ID   DOP2_DROME     STANDARD;      PRT;   539 AA.
AC   Q24563; Q24569; Q9VAJ8;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
```

```
DE   Dopamine receptor 2.
GN   DOPR2 OR DOPR99B OR DAMB OR CG18741.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM LONG).
RC   STRAIN=Canton-S; TISSUE=Head;
RX   MEDLINE=96242088; PubMed=8656286;
RA   Feng G., Hannan F., Reale V., Hon Y.Y., Kousky C.T., Evans P.D.,
RA   Hall L.M.;
RT   "Cloning and functional characterization of a novel dopamine receptor
RT   from Drosophila melanogaster.";
RL   J. Neurosci. 16:3925-3933(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Canton-S;
RX   MEDLINE=96258265; PubMed=8663989;
RA   Han K.-A., Millar N.S., Grotewiel M.S., Davis R.L.;
RT   "DAMB, a novel dopamine receptor expressed specifically in Drosophila
RT   mushroom bodies.";
RL   Neuron 16:1127-1135(1996).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: Receptor for dopamine. The activity of this receptor is
CC       mediated by G proteins which activate adenylyl cyclase. Also
CC       capable of generating a calcium signal. In terms of antagonist
CC       responses, would be classed with the D1-like dopamine receptor
CC       group. This receptor an attractive candidate for initiating
CC       biochemical cascades underlying olfactory learning.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=Q24563-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=Q24563-2; Sequence=VSP_001877;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Expressed in both central and peripheral
CC       nervous systems.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34383; AAC47161.1; -.
DR   EMBL; U61264; AAB08000.1; -.
DR   EMBL; AE003770; AAF56908.2; -.
DR   FlyBase; FBgn0015129; DopR2.
DR   GO; GO:0004952; F:dopamine receptor activity; IDA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PRINTS; PR00960; LMBPPROTEIN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation; Lipoprotein; Palmitate; Alternative splicing.
FT   DOMAIN        1    113       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    114    134       1 (POTENTIAL).
FT   DOMAIN      135    145       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    146    166       2 (POTENTIAL).
FT   DOMAIN      167    189       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    190    206       3 (POTENTIAL).
FT   DOMAIN      207    227       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    228    248       4 (POTENTIAL).
FT   DOMAIN      249    266       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    267    287       5 (POTENTIAL).
FT   DOMAIN      288    420       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    421    441       6 (POTENTIAL).
FT   DOMAIN      442    453       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    454    474       7 (POTENTIAL).
FT   DOMAIN      475    539       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       86     89       POLY-GLY.
FT   DOMAIN      338    351       POLY-GLY.
FT   DOMAIN      357    367       POLY-HIS.
FT   CARBOHYD      5      5       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     31     31       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    182    261       BY SIMILARITY.
FT   LIPID       492    492       S-palmitoyl cysteine (Potential).
FT   LIPID       493    493       S-palmitoyl cysteine (Potential).
FT   VARSPLIC    511    539       RFATRRCYSTCSLHGIQHVRHNSCEQTYI -> CHVAAAMV
FT                                AASTSFGYHSVNQIDRTLM (in isoform Short).
FT                                /FTId=VSP_001877.
SQ   SEQUENCE   539 AA;  59505 MW;  32FDDC0E935AF4B3 CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 539;
  Best Local Similarity   86.7%;  Pred. No. 0.5;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db        338 GGGGGGGGGGGGGGS 352


RESULT 11
T7L1_HUMAN
```

```
ID   T7L1_HUMAN     STANDARD;      PRT;   588 AA.
AC   Q9HCS4; Q9NP00;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Transcription factor 7-like 1 (HMG-box transcription factor 3) (TCF-
DE   3).
GN   TCF7L1 OR TCF3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal lung;
RX   MEDLINE=20535962; PubMed=11085512;
RA   Sagara N., Katoh M.;
RT   "Mitomycin C resistance induced by TCF-3 overexpression in gastric
RT   cancer cell line MKN28 is associated with DT-diaphorase down-
RT   regulation.";
RL   Cancer Res. 60:5959-5962(2000).
RN   [2]
RP   SEQUENCE OF 331-419 FROM N.A.
RX   MEDLINE=92158676; PubMed=1741298;
RA   Castrop J., van Norren K., Clevers H.C.;
RT   "A gene family of HMG-box transcription factors with homology to TCF-
RT   1.";
RL   Nucleic Acids Res. 20:611-611(1992).
RN   [3]
RP   TISSUE-SPECIFICITY.
RX   MEDLINE=99113953; PubMed=9916915;
RA   Barker N., Huls G., Korinek V., Clevers H.;
RT   "Restricted high level expression of Tcf-4 protein in intestinal and
RT   mammary gland epithelium.";
RL   Am. J. Pathol. 154:29-35(1999).
CC   -!- FUNCTION: Participates in the Wnt signaling pathway. Binds to DNA
CC       and acts as repressor in the absence of CTNNB1, and as activator
CC       in its presence. Necessary for the terminal differentiation of
CC       epidermal cells, the formation of keratohyalin granules and the
CC       development of the barrier function of the epidermis (By
CC       similarity). Down-regulates NQO1, leading to increased mitomycin c
CC       resistance.
CC   -!- SUBUNIT: Binds the armadillo repeat of CTNNB1 and forms a stable
CC       complex (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Detected in hair follicles and skin
CC       keratinocytes, and at lower levels in stomach epithelium.
CC   -!- DOMAIN: The putative Groucho interaction domain between the N-
CC       terminal CTNNB1 binding domain and the HMG-box is necessary for
CC       repression of the transactivation mediated by TCF7L1 and CTNNB1
CC       (By similarity).
CC   -!- SIMILARITY: Belongs to the TCF/LEF family.
CC   -!- SIMILARITY: Contains 1 HMG box domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB031046; BAB18185.1; -.
DR   EMBL; X62870; CAB91064.1; -.
DR   HSSP; P27782; 2LEF.
DR   Genew; HGNC:11640; TCF7L1.
DR   MIM; 604652; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0006325; P:establishment and/or maintenance of chromat. . .; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0030111; P:regulation of Wnt receptor signaling pathway; NAS.
DR   InterPro; IPR000910; HMG_12_box.
DR   Pfam; PF00505; HMG_box; 1.
DR   SMART; SM00398; HMG; 1.
DR   PROSITE; PS50118; HMG_BOX_2; 1.
KW   Transcription regulation; Activator; Repressor; Trans-acting factor;
KW   Nuclear protein; DNA-binding; Wnt signaling pathway.
FT   DOMAIN        1     74       CTNNB1-BINDING (BY SIMILARITY).
FT   DNA_BIND    346    414       HMG BOX.
FT   DOMAIN      421    427       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN        5     29       GLY-RICH.
FT   DOMAIN      117    326       PRO-RICH.
SQ   SEQUENCE   588 AA;  62630 MW;  82FB0C9300482A02 CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 588;
  Best Local Similarity   92.9%;  Pred. No. 0.54;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db          6 GGGGGGGGGSGGGG 19


RESULT 12
SR68_CANFA
ID   SR68_CANFA     STANDARD;      PRT;   622 AA.
AC   Q00004;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Signal recognition particle 68 kDa protein (SRP68).
GN   SRP68.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=91092392; PubMed=1702390;
RA   Herz J., Flint N., Stanley K., Frank R., Dobberstein B.;
RT   "The 68 kDa protein of signal recognition particle contains a
RT   glycine-rich region also found in certain RNA-binding proteins.";
RL   FEBS Lett. 276:103-107(1990).
CC   -!- FUNCTION: Signal-recognition-particle assembly has a crucial role
CC       in targeting secretory proteins to the rough endoplasmic
CC       reticulum membrane. SRP68 binds the 7S RNA, SRP72 binds to this
CC       complex subsequently. This ribonucleoprotein complex might
CC       interact directly with the docking protein in the ER membrane and
CC       possibly participate in the elongation arrest function.
CC   -!- SUBUNIT: Signal recognition particle consists of a 7S RNA molecule
CC       of 300 nucleotides and six protein subunits: SRP72, SRP68, SRP54,
CC       SRP19, SRP14 and SRP9.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear; nucleolar (By
CC       similarity).
CC   -!- MISCELLANEOUS: The RNA binding domain is located near the N-
CC       terminus.
CC   -!- SIMILARITY: Belongs to the SRP68 family.
CC   -!- CAUTION: Some authors found genomic clones that have 9 or 12
CC       consecutive glycine residues instead of 15 (AA 9-27).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53744; CAA37773.1; ALT_SEQ.
DR   PIR; A58947; A58947.
DR   InterPro; IPR008941; TPR-like.
KW   Signal recognition particle; Ribonucleoprotein; RNA-binding;
KW   Nuclear protein.
FT   DOMAIN        9     27       POLY-GLY.
SQ   SEQUENCE   622 AA;  70275 MW;  DB03DFB0DAE8B942 CRC64;
```

```
  Query Match             90.5%;  Score 76;  DB 1;  Length 622;
  Best Local Similarity   86.7%;  Pred. No. 0.57;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db         10 GGGGGGGGGGGGGGS 24


RESULT 13
DUS8_MOUSE
ID   DUS8_MOUSE      STANDARD;      PRT;   663 AA.
AC   O09112;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Dual specificity protein phosphatase 8 (EC 3.1.3.48) (EC 3.1.3.16)
DE   (Neuronal tyrosine threonine phosphatase 1).
GN   DUSP8 OR NTTP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96311565; PubMed=8733137;
RA   Theodosiou A.M., Rodrigues N.R., Nesbit M.A., Ambrose H.J.,
RA   Paterson H., McLellan-Arnold E., Boyd Y., Leversha M.A., Owen N.,
RA   Blake D.J., Ashworth A., Davies K.E.;
RT   "A member of the MAP kinase phosphatase gene family in mouse
RT   containing a complex trinucleotide repeat in the coding region.";
RL   Hum. Mol. Genet. 5:675-684(1996).
CC   -!- FUNCTION: This protein shows both activity toward tyrosine-protein
CC       phosphate as well as with serine/threonine-protein phosphate (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
CC       phosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in brain and lung.
CC   -!- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC       Non-receptor class dual specificity subfamily.
CC   -!- SIMILARITY: Contains 1 rhodanese domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X95518; CAA64772.1; -.
DR   HSSP; Q16828; 1MKP.
DR   MGD; MGI:106626; Dusp8.
DR   InterPro; IPR000340; DS_phosphatase.
DR   InterPro; IPR008343; MAPK_phosph.
DR   InterPro; IPR001763; Rhodanese-like.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   Pfam; PF00782; DSPc; 1.
DR   PRINTS; PR01764; MAPKPHPHTASE.
DR   SMART; SM00195; DSPc; 1.
DR   SMART; SM00450; RHOD; 1.
DR   PROSITE; PS50206; RHODANESE_3; 1.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
KW   Hydrolase; Nuclear protein.
FT   DOMAIN       23    138       RHODANESE.
FT   DOMAIN      162    432       PROTEIN-TYROSINE PHOSPHATASE.
FT   DOMAIN      452    459       POLY-ARG.
FT   DOMAIN      555    558       POLY-SER.
FT   DOMAIN      559    576       POLY-GLY.
FT   DOMAIN      577    600       POLY-SER.
FT   DOMAIN      311    552       PRO-RICH.
FT   ACT_SITE    246    246       PHOSPHOCYSTEINE INTERMEDIATE (BY
FT                                SIMILARITY).
SQ   SEQUENCE   663 AA;  68847 MW;  416F429A12C1FA7C CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 663;
  Best Local Similarity   86.7%;  Pred. No. 0.6;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db        563 GGGGGGGGGGGGGGS 577


RESULT 14
EOMD_MOUSE
ID   EOMD_MOUSE      STANDARD;      PRT;   688 AA.
AC   O54839; Q9QYG7;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Eomesodermin homolog.
GN   EOMES OR TBR2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99337662; PubMed=10407135;
RA   Kimura N., Nakashima K., Ueno M., Taga T.;
RT   "A novel mammalian T-box-containing gene, Tbr2, expressed in mouse
RT   developing brain.";
RL   Brain Res. Dev. Brain Res. 115:183-193(1999).
RN   [2]
RP   SEQUENCE OF 278-457 FROM N.A.
RX   MEDLINE=98163742; PubMed=9503012;
RA   Wattler S., Russ A., Evans M., Nehls M.;
RT   "A combined analysis of genomic and primary protein structure defines
RT   the phylogenetic relationship of new members if the T-box family.";
RL   Genomics 48:24-33(1998).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: Contains 1 T-box domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB031037; BAA83416.1; -.
DR   EMBL; AF013281; AAC16233.1; -.
DR   HSSP; P24781; 1XBR.
DR   MGD; MGI:1201683; Eomes.
DR   InterPro; IPR008967; P53-like.
DR   InterPro; IPR001699; TF_T-box.
DR   Pfam; PF00907; T-box; 1.
DR   PRINTS; PR00937; TBOX.
DR   SMART; SM00425; TBOX; 1.
DR   PROSITE; PS01283; TBOX_1; 1.
DR   PROSITE; PS01264; TBOX_2; 1.
DR   PROSITE; PS50252; TBOX_3; 1.
KW   Developmental protein; Transcription regulation; DNA-binding;
KW   Nuclear protein.
FT   DOMAIN       27     41       POLY-GLY.
FT   DNA_BIND    278    458       T-BOX.
```

```
FT   DOMAIN      383    386       POLY-ASN.
SQ   SEQUENCE   688 AA;  72638 MW;  197B0B89E920B82B CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 688;
  Best Local Similarity   86.7%;  Pred. No. 0.62;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db         28 GGGGGGGGGGGGGGS 42


RESULT 15
RFX1_HUMAN
ID   RFX1_HUMAN     STANDARD;      PRT;   979 AA.
AC   P22670;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   MHC class II regulatory factor RFX1 (RFX) (Enhancer factor C) (EF-C).
GN   RFX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91071581; PubMed=2253877;
RA   Reith W., Sanchez-Herrero C., Kobr M., Silacci P., Berte C.,
RA   Barras E., Mach B.;
RT   "MHC class II regulatory factor RFX has a novel DNA-binding domain
RT   and a functionally independent dimerization domain.";
RL   Genes Dev. 4:1528-1540(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   IDENTITY BETWEEN RFX1 AND EF-C.
RX   MEDLINE=94019311; PubMed=8413236;
RA   Siegrist C.A., Durand B., Emery P., David E., Hearing P., Mach B.,
RA   Reith W.;
RT   "RFX1 is identical to enhancer factor C and functions as a
RT   transactivator of the hepatitis B virus enhancer.";
RL   Mol. Cell. Biol. 13:6375-6384(1993).
RN   [4]
RP   BINDING TO RPL30 PROMOTER.
RX   MEDLINE=94040774; PubMed=8224874;
RA   Safrany G., Perry R.P.;
RT   "Transcription factor RFX1 helps control the promoter of the mouse
RT   ribosomal protein-encoding gene rpL30 by binding to its alpha
RT   element.";
RL   Gene 132:279-283(1993).
RN   [5]
RP   SHOWS THAT BLS II IS NOT DUE TO RFX1.
RX   MEDLINE=92375076; PubMed=1508204;
RA   Sanchez-Herrero C., Reith W., Silacci P., Mach B.;
RT   "The DNA-binding defect observed in major histocompatibility complex
RT   class II regulatory mutants concerns only one member of a family of
RT   complexes binding to the X boxes of class II promoters.";
RL   Mol. Cell. Biol. 12:4076-4083(1992).
CC   -!- FUNCTION: Regulatory factor essential for MHC class II genes
CC       expression. Binds to the X boxes of MHC class II genes. Also binds
CC       to an inverted repeat (ENH1) required for hepatitis B virus genes
CC       expression and to the most upstream element (alpha) of the RPL30
CC       promoter.
CC   -!- SUBUNIT: Binds DNA as a homodimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the RFX family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X58964; CAA41730.1; -.
DR   EMBL; A20498; CAA01506.1; -.
DR   EMBL; BC049826; AAH49826.1; -.
DR   PIR; A35913; A35913.
DR   PDB; 1DP7; 06-MAR-00.
DR   TRANSFAC; T01673; -.
DR   Genew; HGNC:9982; RFX1.
DR   MIM; 600006; -.
DR   GO; GO:0003705; F:RNA polymerase II transcription factor acti. . .; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   InterPro; IPR007668; RFX1_trans_act.
DR   InterPro; IPR003150; RFX_DNA_binding.
DR   Pfam; PF04589; RFX1_trans_act; 1.
DR   Pfam; PF02257; RFX_DNA_binding; 1.
KW   DNA-binding; Transcription regulation; Activator; Nuclear protein;
KW   3D-structure.
FT   DOMAIN      381    411       GLY-RICH.
FT   DNA_BIND    438    528       EXPERIMENTALLY DEDUCED.
FT   DOMAIN      920    936       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      744    979       NECESSARY FOR DIMERIZATION.
SQ   SEQUENCE   979 AA;  104728 MW;  556151F88C6AC9A2 CRC64;

  Query Match             90.5%;  Score 76;  DB 1;  Length 979;
  Best Local Similarity   92.9%;  Pred. No. 0.84;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              ||||||||| ||||
Db        381 GGGGSGGGGGGGGG 394


Search completed: April 20, 2004, 10:25:36
Job time : 14.6 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:22:14 ; Search time 16.2 Seconds
(without alignments)
89.066 Million cell updates/sec

Title: US-08-930-480A-5
Perfect score: 84
Sequence: 1 GGGGSGGGGSGGGGS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	80	95.2	592	2	E82759	endo-1,4-beta-gluc
2	79	94.0	104	2	JC4190	holotricin 3 precu
3	78	92.9	249	2	S41374	single chain Fv an
4	77	91.7	206	2	I53066	gene M-twist prote
5	77	91.7	268	2	A56446	Ig heavy chain V r
6	76	90.5	80	2	T10550	hypothetical prote
7	76	90.5	165	1	KNRZG1	glycine-rich cell
8	76	90.5	207	2	T07381	glycine-rich prote
9	76	90.5	221	2	T04592	glycine-rich cell
10	76	90.5	255	2	B84777	hypothetical prote
11	76	90.5	280	2	A42424	chitinase (EC 3.2.
12	76	90.5	322	2	A45036	single-stranded-DN
13	76	90.5	378	2	S04336	U1 snRNP 70K prote
14	76	90.5	433	2	S20963	homeotic protein H
15	76	90.5	528	2	G02127	fus-like protein -
16	76	90.5	979	2	A35913	regulatory factor
17	76	90.5	1969	2	T08875	histidine kinase h
18	76	90.5	2783	1	A41948	alpha-fetoprotein
19	75	89.3	302	2	C84470	hypothetical prote
20	75	89.3	322	2	T04595	hypothetical prote
21	75	89.3	1226	2	T24045	hypothetical prote
22	74	88.1	266	1	CIRBL	calpain (EC 3.4.22
23	74	88.1	440	2	S71795	transcription fact
24	74	88.1	722	1	VCPVV2	coat protein VP1 -
25	74	88.1	877	2	T43449	hypothetical prote
26	74	88.1	895	2	JC7089	zinc finger bindin
27	73	86.9	151	2	S43296	bone morphogenetic
28	73	86.9	211	2	T04098	CBP20 preproprotei
29	73	86.9	263	2	A34466	calpain (EC 3.4.22
30	73	86.9	266	1	CIPGL	calpain (EC 3.4.22
31	73	86.9	268	1	CIHUL	calpain (EC 3.4.22
32	73	86.9	272	2	T02745	nucleic acid bindi
33	73	86.9	291	1	S31415	glycine-rich prote
34	73	86.9	333	2	A39065	homeotic protein E
35	73	86.9	367	2	JC6087	helix-loop-helix t
36	73	86.9	377	2	T04213	heat shock transcr
37	73	86.9	396	2	T49109	glycine-rich prote
38	73	86.9	495	1	S31223	transcription fact
39	73	86.9	681	2	AB2155	hypothetical prote
40	73	86.9	901	2	JC6093	dead ringer nuclea
41	73	86.9	1473	2	T13855	suppressor of sabl
42	72	85.7	106	2	F84797	hypothetical prote
43	72	85.7	113	2	S44750	C06G4.3 protein -
44	72	85.7	136	2	T29282	hypothetical prote
45	72	85.7	188	2	S49192	GCR 1 protein - fr

ALIGNMENTS

RESULT 1
E82759
endo-1,4-beta-glucanase XF0818 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: E82759
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: E82759
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <SIM>
A;Cross-references: GB:AE003921; GB:AE003849; NID:g9105710; PIDN:AAF83628.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0818

Query Match 95.2%; Score 80; DB 2; Length 592;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGGSGGGGSGGGG 14
              ||||||||||||||
Db        467 GGGGSGGGGSGGGG 480
```

RESULT 2
JC4190
holotricin 3 precursor - Holotrichia diomphalia
N;Alternate names: antifungal protein
C;Species: Holotrichia diomphalia
C;Date: 04-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C;Accession: JC4190
R;Lee, S.Y.; Moon, H.J.; Kurata, S.; Natori, S.; Lee, B.L.
Biol. Pharm. Bull. 18, 1049-1052, 1995

```
A;Title: Purification and cDNA cloning of an antifungal protein from the hemolymph of Hc
A;Reference number: JC4190; MUID:96073722; PMID:8535393
A;Accession: JC4190
A;Molecule type: mRNA
A;Residues: 1-104 <LEE>
A;Cross-references: DDBJ:D13744; NID:g1088433; PIDN:BAA02889.1; PID:d1003394; PID:g17861
C;Comment: This protein is a Gly- and His-rich protein and a constitutive protein of lar
C;Keywords: hemolymph
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-104/Product: holotricin 3 #status predicted <MAT>

  Query Match             94.0%;  Score 79;  DB 2;  Length 104;
  Best Local Similarity   93.3%;  Pred. No. 0.08;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| ||||||||||
Db         64 GGGGRGGGGSGGGGS 78


RESULT 3
S41374
single chain Fv antibody - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995
C;Accession: S41374
R;Artsaenko, O.; Weiler, E.W.; Muentz, K.; Conrad, U.
submitted to the EMBL Data Library, January 1994
A;Description: Construction and functional characterization of a single chain Fv antibod
A;Reference number: S41374
A;Accession: S41374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-249 <ART>
A;Cross-references: EMBL:Z29480

  Query Match             92.9%;  Score 78;  DB 2;  Length 249;
  Best Local Similarity   100.0%;  Pred. No. 0.21;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GGGSGGGGSGGGGS 15
              ||||||||||||||
Db        121 GGGSGGGGSGGGGS 134


RESULT 4
I53066
gene M-twist protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I53066; I66795
R;Wolf, C.; Thisse, C.; Stoetzel, C.; Thisse, B.; Gerlinger, P.; Perrin-Schmitt, F.
Dev. Biol. 143, 363-373, 1991
A;Title: The M-twist gene of Mus is expressed in subsets of mesodermal cells and is clos
A;Reference number: I53066; MUID:91122450; PMID:1840517
A;Accession: I53066
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-206 <RES>
A;Cross-references: GB:M63649; NID:g202243; PIDN:AAA40514.1; PID:g202244
A;Accession: I66795
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-35,'R',37-90,'P',92-206 <RE2>
A;Cross-references: GB:M63650; NID:g202245; PIDN:AAA40515.1; PID:g202246
C;Genetics:
A;Gene: M-twist

  Query Match             91.7%;  Score 77;  DB 2;  Length 206;
  Best Local Similarity   86.7%;  Pred. No. 0.22;
  Matches   13;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 15
              ||||:|||| |||||
Db         83 GGGGAGGGGGGGGG 97


RESULT 5
A56446
Ig heavy chain V region (3H-3H scFv) - mouse (strain BALB/C)
C;Species: Mus musculus (house mouse)
C;Date: 19-Jan-1996 #sequence_revision 19-Jan-1996 #text_change 16-Aug-1996
C;Accession: A56446
R;Tang, P.M.; Foltz, L.A.; Mahoney, W.C.; Schueler, P.A.
J. Biol. Chem. 270, 7829-7835, 1995
A;Title: A high affinity digoxin-binding protein displayed on M13 is functionally identi
A;Reference number: A56446; MUID:95229583; PMID:7713873
A;Accession: A56446
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-268 <TAN>
A;Cross-references: GB:U20617
C;Keywords: heterotetramer; immunoglobulin

  Query Match             91.7%;  Score 77;  DB 2;  Length 268;
  Best Local Similarity   93.3%;  Pred. No. 0.28;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||| ||||||
Db        121 GGGGSGGGDSGGGGS 135


RESULT 6
T10550
hypothetical protein T12G13.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C;Accession: T10550
R;Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16533
A;Accession: T10550
A;Molecule type: DNA
A;Residues: 1-80 <BEV>
A;Cross-references: EMBL:AL080252; GSPDB:GN00062; ATSP:T12G13.70
A;Experimental source: cultivar Columbia; BAC clone T12G13
C;Genetics:
A;Gene: ATSP:T12G13.70
A;Map position: 4

  Query Match             90.5%;  Score 76;  DB 2;  Length 80;
  Best Local Similarity   92.9%;  Pred. No. 0.12;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         22 GGGGGGGGGSGGGG 35


RESULT 7
KNRZG1
glycine-rich cell wall structural protein 1 precursor (clone lambda-313) - rice
C;Species: Oryza sativa (rice)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C;Accession: S13385
R;Lei, M.; Wu, R.
Plant Mol. Biol. 16, 187-198, 1991
A;Title: A novel glycine-rich cell wall protein gene in rice.
A;Reference number: S13385; MUID:91370862; PMID:1716496
A;Accession: S13385
A;Molecule type: DNA
A;Residues: 1-165 <LEI>
A;Cross-references: EMBL:X53596; NID:g20246; PIDN:CAA37665.1; PID:g20247
C;Genetics:
A;Gene: grp-1
```

```
C;Superfamily: glycine-rich cell wall structural protein 1
C;Keywords: cell wall; duplication; structural protein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-165/Product: glycine-rich cell wall structural protein 1 #status predicted <MAT>
F;30-55/Region: repeat R1
F;56-62/Region: repeat R2
F;62-92/Region: repeat R1
F;93-99/Region: repeat R2
F;100-131/Region: repeat R1
F;132-138/Region: repeat R2
F;139-160/Region: repeat R1

  Query Match             90.5%;  Score 76;  DB 1;  Length 165;
  Best Local Similarity   86.7%;  Pred. No. 0.23;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db        111 GGGGGGGGGGGGGGS 125


RESULT 8
T07381
glycine-rich protein Tfm5 - tomato
C;Species: Lycopersicon esculentum (tomato)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C;Accession: T07381
R;Santino, C.G.; Stanford, G.L.; Conner, T.W.
Plant Mol. Biol. 33, 405-416, 1997
A;Title: Developmental and transgenic analysis of two tomato fruit enhanced genes.
A;Reference number: Z16000; MUID:97201476; PMID:9049262
A;Accession: T07381
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-207 <SAN>
A;Cross-references: EMBL:X95262; NID:g1166449; PIDN:CAA64559.1; PID:g1166450
A;Experimental source: cultivar UC82b; fruit
C;Genetics:
A;Gene: Tfm5
C;Superfamily: hydroxyproline-rich glycoprotein

  Query Match             90.5%;  Score 76;  DB 2;  Length 207;
  Best Local Similarity   92.9%;  Pred. No. 0.28;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              ||||||||| ||||
Db         97 GGGGSGGGGGGGGG 110


RESULT 9
T04592
glycine-rich cell wall structural protein homolog F23E13.120 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04592
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.
submitted to the Protein Sequence Database, March 1998
A;Reference number: Z15378
A;Accession: T04592
A;Molecule type: DNA
A;Residues: 1-221 <BEV>
A;Cross-references: EMBL:AL022141
A;Experimental source: cultivar Columbia; BAC clone F23E13
C;Genetics:
A;Map position: 4
A;Note: F23E13.120

  Query Match             90.5%;  Score 76;  DB 2;  Length 221;
  Best Local Similarity   86.7%;  Pred. No. 0.3;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db        168 GGGGGGGGGGGGGGS 182


RESULT 10
B84777
hypothetical protein At2g36120 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: B84777
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; F
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-255 <STO>
A;Cross-references: GB:AE002093; NID:g4678224; PIDN:AAD26969.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g36120
A;Map position: 2
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology

  Query Match             90.5%;  Score 76;  DB 2;  Length 255;
  Best Local Similarity   92.9%;  Pred. No. 0.34;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db        213 GGGGGGGGGSGGGG 226


RESULT 11
A42424
chitinase (EC 3.2.1.14) A - maize
C;Species: Zea mays (maize)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-1999
C;Accession: A42424; A42260
R;Huynh, Q.K.; Hironaka, C.M.; Levine, E.B.; Smith, C.E.; Borgmeyer, J.R.; Shah, D.M.
J. Biol. Chem. 267, 6635-6640, 1992
A;Title: Antifungal proteins from plants. Purification, molecular cloning, and antifungal
A;Reference number: A42424; MUID:92202208; PMID:1551872
A;Accession: A42424
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-280 <HUY>
A;Experimental source: seed
A;Note: sequence inconsistent with nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:89874, NCBIP:89876)
R;Verburg, J.G.; Smith, C.E.; Lisek, C.A.; Huynh, Q.K.
J. Biol. Chem. 267, 3886-3893, 1992
A;Title: Identification of an essential tyrosine residue in the catalytic site of a chit
opyl)- carbodiimide.
A;Reference number: A42260; MUID:92156129; PMID:1740436
A;Accession: A42260
A;Molecule type: protein
A;Residues: 180-195 <VER>
A;Note: the residue designated 'X' was determined to be derivatized tyrosine; therefore,
C;Superfamily: lectin-related plant chitinase; hevein chitin-binding domain homology; pla
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;26-61/Domain: hevein chitin-binding domain homology <HCB>
F;82-280/Domain: plant chitinase homology <PCH>
F;188/Active site: Tyr #status predicted

  Query Match             90.5%;  Score 76;  DB 2;  Length 280;
  Best Local Similarity   86.7%;  Pred. No. 0.36;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
```

```
Db         61 GGGGGGGGGGGGGGS 75


RESULT 12
A45036
single-stranded-DNA-binding protein Pur alpha - human
C;Species: Homo sapiens (man)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C;Accession: A45036
R;Bergemann, A.D.; Ma, Z.W.; Johnson, E.M.
Mol. Cell. Biol. 12, 5673-5682, 1992
A;Title: Sequence of cDNA comprising the human pur gene and sequence-specific single-str
A;Reference number: A45036; MUID:93078769; PMID:1448097
A;Accession: A45036
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-322 <BER>
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:119216)

  Query Match             90.5%;  Score 76;  DB 2;  Length 322;
  Best Local Similarity   92.9%;  Pred. No. 0.41;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db         34 GGGGGGGGGSGGGG 47


RESULT 13
S04336
U1 snRNP 70K protein (long form) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 28-Feb-1990 #sequence_revision 30-Sep-1991 #text_change 24-Sep-1999
C;Accession: S04336
R;Hornig, H.; Fischer, U.; Costas, M.; Rauh, A.; Luehrmann, R.
Eur. J. Biochem. 182, 45-50, 1989
A;Title: Analysis of genomic clones of the murine U1RNA-associated 70-kDa protein reveal
A;Reference number: S04336; MUID:89276388; PMID:2525092
A;Accession: S04336
A;Molecule type: DNA
A;Residues: 1-378 <HOR>
A;Cross-references: EMBL:X15769; NID:g55084; PIDN:CAA33777.1; PID:g763157
A;Note: the authors translated the codon GGC for residue 101 as Glu, ACT for residue 113
es 368 and 374 as Asp
C;Genetics:
A;Introns: 19/1; 40/3; 61/3; 89/1; 123/1; 152/2
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotei
C;Keywords: alternative splicing
F;34-101/Domain: ribonucleoprotein repeat homology <RRM>

  Query Match             90.5%;  Score 76;  DB 2;  Length 378;
  Best Local Similarity   92.9%;  Pred. No. 0.47;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db        241 GGGGGGGGGSGGGG 254


RESULT 14
S20963
homeotic protein Hox B3 - mouse
N;Alternate names: homeotic protein Hox 2.7
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 21-Jul-1995 #text_change 20-Aug-1999
C;Accession: S20963; D42694
R;Sham, M.H.; Hunt, P.; Nonchev, S.; Papalopulu, N.; Graham, A.; Boncinelli, E.; Krumlau
EMBO J. 11, 1825-1836, 1992
A;Title: Analysis of the murine Hox-2.7 gene: conserved alternative transcripts with dif
A;Reference number: S20963; MUID:92258392; PMID:1582411
A;Accession: S20963
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-433 <SHA>
A;Cross-references: GB:X66177; GB:S35628; GB:S35738; NID:g312229; PIDN:CAA46951.1; PID:g
R;Nazarali, A.; Kim, Y.; Nirenberg, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 2883-2887, 1992
A;Title: Hox-1.11 and Hox-4.9 homeobox genes.
A;Reference number: A42694; MUID:92212934; PMID:1348361
A;Accession: D42694
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 213-238 <NAZ>
A;Note: sequence extracted from NCBI backbone (NCBIN:92310, NCBIP:92316)
C;Superfamily: homeotic protein Hox B3; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;192-248/Domain: homeobox homology <HOX>

  Query Match             90.5%;  Score 76;  DB 2;  Length 433;
  Best Local Similarity   86.7%;  Pred. No. 0.53;
  Matches   13;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||| |||| |||||
Db        156 GGGGGGGGGGGGGGS 170


RESULT 15
G02127
fus-like protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 28-Jul-2003
C;Accession: G02127
R;Itoh, K.; Kawase, M.
submitted to the EMBL Data Library, September 1995
A;Reference number: G09199
A;Accession: G02127
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-528 <ITO>
A;Cross-references: EMBL:U36561; NID:g1040969; PIDN:AAA79948.1; PID:g1040970
C;Superfamily: RNA-binding protein, EWS type; ribonucleoprotein repeat homology
F;289-364/Domain: ribonucleoprotein repeat homology <RRM>

  Query Match             90.5%;  Score 76;  DB 2;  Length 528;
  Best Local Similarity   92.9%;  Pred. No. 0.63;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGG 14
              |||| |||||||||
Db        183 GGGGGGGGGSGGGG 196


Search completed: April 20, 2004, 10:27:41
Job time : 17.2 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 10:20:04 ; Search time 60.6 Seconds
(without alignments)
69.938 Million cell updates/sec

Title: US-08-930-480A-5
Perfect score: 84
Sequence: 1 GGGGSGGGGSGGGGS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	84	100.0	15	2	AAR25983	Aar25983 Peptide m
2	84	100.0	15	2	AAR59500	Aar59500 Hydrophil
3	84	100.0	15	2	AAR85123	Aar85123 Gene deli
4	84	100.0	15	2	AAR76683	Aar76683 Human ONS
5	84	100.0	15	2	AAR99244	Aar99244 (Gly4Ser)
6	84	100.0	15	2	AAR95067	Aar95067 scFv spac
7	84	100.0	15	2	AAW09323	Aaw09323 Peptide l
8	84	100.0	15	2	AAY49219	Aay49219 Sequence
9	84	100.0	15	2	AAW10295	Aaw10295 Peptide l
10	84	100.0	15	2	AAW35984	Aaw35984 Peptide l
11	84	100.0	15	2	AAW87784	Aaw87784 Antibody-
12	84	100.0	15	2	AAY43414	Aay43414 Peptide S
13	84	100.0	15	2	AAY33328	Aay33328 E6-sFv pe
14	84	100.0	15	2	AAY03763	Aay03763 Linker pe
15	84	100.0	15	2	AAY21600	Aay21600 EP-919566
16	84	100.0	15	2	AAY27397	Aay27397 Flexible
17	84	100.0	15	2	AAE16564	Aae16564 Peptide l
18	84	100.0	15	3	AAB29542	Aab29542 Linker pe
19	84	100.0	15	3	AAY99636	Aay99636 (gly4ser)
20	84	100.0	15	3	AAB22838	Aab22838 Single ch
21	84	100.0	15	3	AAY70606	Aay70606 Protein e
22	84	100.0	15	3	AAY79551	Aay79551 Linker pe
23	84	100.0	15	3	AAY79552	Aay79552 Linker pe
24	84	100.0	15	3	AAY90826	Aay90826 Linker am
25	84	100.0	15	3	AAB23816	Aab23816 Plasmid p
26	84	100.0	15	3	AAY97237	Aay97237 Peptide l
27	84	100.0	15	3	AAB15682	Aab15682 Single-ch
28	84	100.0	15	4	AAB70169	Aab70169 Gly/Ser l
29	84	100.0	15	4	AAB98920	Aab98920 Linker pe
30	84	100.0	15	4	AAE13082	Aae13082 Glycine l
31	84	100.0	15	4	AAM52571	Aam52571 Peptide l
32	84	100.0	15	4	AAE12408	Aae12408 Peptide l
33	84	100.0	15	4	AAU08689	Aau08689 Antibody
34	84	100.0	15	4	AAE06268	Aae06268 Glycine l
35	84	100.0	15	4	AAU09970	Aau09970 Glycine-S
36	84	100.0	15	4	AAU04948	Aau04948 Humanised
37	84	100.0	15	4	AAE13100	Aae13100 Linker pe
38	84	100.0	15	4	ABB79010	Abb79010 Peptide l
39	84	100.0	15	4	AAB97229	Aab97229 Immunoglo
40	84	100.0	15	4	AAB85296	Aab85296 Sequence
41	84	100.0	15	4	AAU29009	Aau29009 Tumour-sp
42	84	100.0	15	4	AAB58601	Aab58601 Peptide e
43	84	100.0	15	4	AAB48032	Aab48032 Gly-Ser p
44	84	100.0	15	4	AAB61572	Aab61572 Flexible
45	84	100.0	15	4	AAB74579	Aab74579 Context-d

ALIGNMENTS

RESULT 1
AAR25983
ID AAR25983 standard; protein; 15 AA.
XX
AC AAR25983;
XX
DT 25-MAR-2003 (revised)
DT 21-JAN-1993 (first entry)
XX
DE Peptide monomer 21.
XX
KW Reverse peptide; microbial pathogen; phytotoxicity; head-to-tail;
KW proteolytic degradation; dimer; peptide bond; bridging group; omega loop.
XX
OS Synthetic.
XX
PN EP497366-A2.
XX
PD 05-AUG-1992.
XX
PF 31-JAN-1992; 92EP-00101616.
XX
PR 01-FEB-1991; 91US-00649784.
XX
PA (DONG) IST DONEGANI SPA GUIDO.
PA (ENIE) ENICHEM SPA.
XX
PI Mapelli C, Dugas De Robertis C, Stahl GF, Bascomb NF;
PI Swerdloff MD, Williams JI, Everett NP;
XX
DR WPI; 1992-260816/32.
XX
PT Reverse antimicrobial peptide(s) and oligopeptide(s) - useful for
PT protecting plants from pathogens and for determining phytotoxicity.
XX
PS Disclosure; Fig 1; 79pp; English.
XX
CC The sequences given in AAR25963-83 are a collection of natural and
CC reverse peptides which are active against at least one microbial pathogen
CC and, preferably, at least one plant pathogen. It has been found that
CC acceptable activity and acceptable levels of protection against at least
CC one microbial pathogen and at least one microbial plant pathogen may be
CC obtained by reversing the sequence of amino acids contained within
CC naturally occuring antimicrobial peptides while maintaining the
CC directionality of the peptide bonds. These peptides possess relatively
CC low phytotoxicity and/or low susceptibility to proteolytic degradation.
CC The oligopeptides may be used as dimers composed of two peptide units
CC with or without an intervening bridge. The simplest structure taken by

```
CC   these dimers is the "head-to-tail" configuration. This comprises at least
CC   one first peptide monomer and at least one second peptide monomer. Each
CC   peptide monomer has an N- and C-terminus, both of which are capable of
CC   forming peptide bonds. In the head-to-tail configuration the C-terminal
CC   amino acid of the first monomer peptide is directly bound to the N-
CC   terminus of the second monomer peptide, by a peptide bond, without an
CC   intervening bridging group. In other peptide dimers bridging groups may
CC   be used and may be as few as one amino acid but may be as large as 100
CC   amino acids in length and form omega loops or other secondary structures.
CC   (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
CC   correct PA field.)
XX
SQ   Sequence 15 AA;

  Query Match             100.0%;  Score 84;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db          1 GGGGSGGGGSGGGGS 15


RESULT 2
AAR59500
ID   AAR59500 standard; peptide; 15 AA.
XX
AC   AAR59500;
XX
DT   25-MAR-2003  (revised)
DT   29-JUL-1994  (first entry)
XX
DE   Hydrophilic linker #1 to make single chain antibody.
XX
KW   Single chain antibody; sFv; heavy chain; light chain; variable domain;
KW   hydrophilic linker; antibodies.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .5
FT                   /note= "first of 3 repeat units"
XX
PN   WO9402610-A1.
XX
PD   03-FEB-1994.
XX
PF   16-JUL-1993;   93WO-US006735.
XX
PR   17-JUL-1992;   92US-00916939.
PR   17-MAR-1993;   93US-00045274.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
XX
PI   Marasco WA,  Haseltine WA;
XX
DR   WPI; 1994-048868/06.
XX
PT   Intracellular binding of antigens - by using antibody targetting with
PT   vector system, for e.g. tumour suppression.
XX
PS   Claim 35; Page 25; 155pp; English.
XX
CC   New vector systems comprise a sequence adapted for intracellular delivery
CC   and expression contg. a promoter operably linked to an antibody gene
CC   encoding an antibody which binds to a specific target antigen. The
CC   antibody is esp. a single chain antibody in which the heavy and light
CC   chain variable regions are joined via a hydrophilic linker peptide.
CC   Examples of suitable linkers are given in AAR59500- AAR59507, with
CC   AAR59500 being the most preferred linker. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 15 AA;

  Query Match             100.0%;  Score 84;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db          1 GGGGSGGGGSGGGGS 15


RESULT 3
AAR85123
ID   AAR85123 standard; peptide; 15 AA.
XX
AC   AAR85123;
XX
DT   06-JUN-1996  (first entry)
XX
DE   Gene delivery fusion protein flexon peptide.
XX
KW   Targeted nucleic acid; fusion protein; nucleic acid binding domain;
KW   gene delivery domain; cell; GAL4; interleukin; flexon; linker; primer;
KW   amplification; PCR; S.cerevisiae; gene therapy.
XX
OS   Synthetic.
XX
PN   WO9528494-A1.
XX
PD   26-OCT-1995.
XX
PF   17-APR-1995;   95WO-US004738.
XX
PR   15-APR-1994;   94US-00227858.
PR   19-OCT-1994;   94US-00326460.
XX
PA   (TARG-) TARGETED GENETICS CORP.
XX
PI   Overell RW,  Weisser KE;
XX
DR   WPI; 1995-373808/48.
DR   N-PSDB; AAT02970.
XX
PT   Fusion protein for delivering targetted nucleic acid to target cell -
PT   comprises a nucleic acid binding domain and a gene delivery domain, used
PT   in, e.g. gene therapy of Cystic fibrosis and in tumour vaccines.
XX
PS   Example 3; Page 49; 80pp; English.
XX
CC   A novel method of delivering a targeted nucleic acid involves a fusion
CC   protein comprising nucleic acid binding domain (NBD) linked to a gene
CC   delivery domain (GDD). The NBD binds the target DNA whilst the GDD
CC   mediates the delivery of the target DNA into the cell. An example of the
CC   fusion protein comprises the GAL4 NBD linked to the interleukin (IL)-2
CC   GDD. The NBD and GDD domain can be separated by a short flexible peptide
CC   linker termed a "flexon". The oligomers AAT02970-1 were annealed to
CC   encode such a "flexon". The annealed product was inserted between the
CC   coding sequence of the yeast GAL4 NBD and the IL-2 GDD in the plasmid
CC   pT3GAL4/IL-2m. This vector was transformed into E.coli DH10B for
CC   production of the fusion protein. The fusion protein has applications in
CC   gene therapy esp. for in vivo and in vitro gene delivery
XX
SQ   Sequence 15 AA;

  Query Match             100.0%;  Score 84;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db          1 GGGGSGGGGSGGGGS 15


RESULT 4
```

```
AAR76683
ID   AAR76683 standard; protein; 15 AA.
XX
AC   AAR76683;
XX
DT   18-JAN-1996  (first entry)
XX
DE   Human ONS-M21 antibody Fv fragment linker peptide.
XX
KW   Plasmid pSCFVT7-hM21; human; ONS-M21 antibody; chimeric protein;
KW   medulloblastoma; brain tumour; treatment; diagnosis; Fv fragment.
XX
OS   Synthetic.
XX
PN   WO9514041-A1.
XX
PD   26-MAY-1995.
XX
PF   19-OCT-1994;   94WO-JP001763.
XX
PR   19-NOV-1993;   93JP-00291078.
XX
PA   (CHUS ) CHUGAI SEIYAKU KK.
XX
PI   Ohtomo T,  Sato K,  Tsuchiya M;
XX
DR   WPI; 1995-200347/26.
DR   N-PSDB; AAQ94549.
XX
PT   Reconstituted antibody against human medullo:blastoma cells - contains
PT   high proportion of human antibody origin and has low antigenicity.
XX
PS   Claim 32; Page 103; 120pp; Japanese.
XX
CC   AAQ94549 encodes AAR76683 a peptide linker, part of the human antibody
CC   ONS-M21 Fv fragment. The fragment was used in the construction of a
CC   human/murine chimeric antibody, reactive with human medullo- blastoma (a
CC   brain tumour) cells. The chimeric antibody can be used in the diagnosis
CC   and treatment of this disease
XX
SQ   Sequence 15 AA;

  Query Match             100.0%;  Score 84;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db          1 GGGGSGGGGSGGGGS 15


RESULT 5
AAR99244
ID   AAR99244 standard; peptide; 15 AA.
XX
AC   AAR99244;
XX
DT   28-NOV-1996  (first entry)
XX
DE   (Gly4Ser)3 linker.
XX
KW   Bioactive fusion protein; interleukin-12; IL-12; p35 subunit;
KW   p40 subunit; antitumour; cytokine; tumour; melanoma; fibrosarcoma;
KW   renal cell carcinoma; immunotherapy; therapy; retrovirus; vector.
XX
OS   Synthetic.
XX
PN   WO9624676-A1.
XX
PD   15-AUG-1996.
XX
PF   07-FEB-1996;   96WO-US001787.
XX
PR   08-FEB-1995;   95US-00385335.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Lieschke GJ,  Mulligan RC;
XX
DR   WPI; 1996-384448/38.
DR   N-PSDB; AAT35195, AAT35196, AAT35202, AAT35203.
XX
PT   New DNA encoding fusion protein, esp. contg. IL-12 p35 and p40 subunits -
PT   for treatment of established tumours or prevention of tumour
PT   establishment.
XX
PS   Claim 2; Page 69; 118pp; English.
XX
CC   Peptide linkers (Gly4Ser)2Ser, (Gly4Ser)3Ser, (Gly4Ser)3 and (Gly6)Ser
CC   (AAR99242-45) are used to join the subunits of novel dimeric or
CC   multimeric fusion proteins. They have been utilised in the prodn. of
CC   bioactive interleukin-12 (IL-12) fusion proteins, linking mouse/human IL-
CC   12 p35 subunit (see also AAR99246) to mouse/human IL-12 p40 subunit
CC   (AAR99247). DNA encoding such constructs can be incorporated into a
CC   retroviral vector (see also AAT35198) to allow dimeric IL-12 prodn. in
CC   transfected cells. Tumour cells (esp. CMS-5, B16 or renal carcinoma
CC   cells) secreting IL-12 dimer can be used to reduce the size of
CC   established tumours and/or increase survival time, esp. in cases of
CC   melanoma, fibrosarcoma and renal cell carcinoma
XX
SQ   Sequence 15 AA;

  Query Match             100.0%;  Score 84;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db          1 GGGGSGGGGSGGGGS 15


RESULT 6
AAR95067
ID   AAR95067 standard; peptide; 15 AA.
XX
AC   AAR95067;
XX
DT   18-AUG-1996  (first entry)
XX
DE   scFv spacer peptide.
XX
KW   Nucleic acid transfer system; gene transfer; gene therapy;
KW   cell targeting; multidomain protein; vector; cancer; scFv;
KW   single chain antibody.
XX
OS   Synthetic.
XX
PN   WO9613599-A1.
XX
PD   09-MAY-1996.
XX
PF   31-OCT-1995;   95WO-EP004270.
XX
PR   01-NOV-1994;   94EP-00810627.
XX
PA   (WELS/) WELS W.
XX
PI   Wels W,  Fominaya J;
XX
DR   WPI; 1996-239505/24.
XX
PT   Nucleic acid transfer system for gene therapy, e.g. against cancer -
PT   includes toxin translocation domain to target nucleic acid to specific
PT   cell.
XX
PS   Example 5; Page 29; 106pp; English.
```

XX
CC A spacer peptide (AAR95067) is used to link the light chain variable
CC domain to the heavy chain variable domain of a single chain recombinant
CC antibody (scFv). It allows correct folding of an antigen binding domain
CC present in the variable domains. The scFv is derived from hybridoma FRP5,
CC which produces monoclonal antibody against the HER2 antigen of human
CC tumour cells. It forms the ligand domain of a multidomain protein (see
CC also AAR95053 and AAR95056-58) that is used with an effector nucleic acid
CC in a novel nucleic acid transfer system suitable for gene therapy. The
CC ligand domain has a target cell recognition function and allows cellular
CC internalization of the multidomain protein/nucleic acid complex
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGSGGGGSGGGGS 15
||||||||||||||||
Db 1 GGGGSGGGGSGGGGS 15

RESULT 7
AAW09323
ID AAW09323 standard; peptide; 15 AA.
XX
AC AAW09323;
XX
DT 10-JUN-1997 (first entry)
XX
DE Peptide linker arm #1.
XX
KW Chimaeric; bispecific; DNA binding domain; trans; activator; repressor;
KW diphtheria; Pseudomonas; toxin; thymidine kinase; single chain antibody;
KW pathogen; HIV Tat; papilloma virus; E6/E7; Epstein-Barr virus; EBNA;
KW hyperproliferation; p53; tumour; oligomerisation.
XX
OS Synthetic.
XX
PN WO9630512-A1.
XX
PD 03-OCT-1996.
XX
PF 29-MAR-1996; 96WO-FR000477.
XX
PR 31-MAR-1995; 95FR-00003841.
XX
PA (RHON) RHONE POULENC RORER SA.
XX
PI Bracco L, Schweighoffer F, Tocque B;
XX
DR WPI; 1996-455359/45.
XX
PT Conditional gene expression system triggered by e.g. infection or hyper-
PT proliferation - comprises novel bi:specific proteins having DNA-binding
PT domain and second domain specific for trans-activator or repressor, for
PT gene therapy.
XX
PS Claim 23; Page 45; 81pp; French.
XX
CC The invention relates to novel chimaeric, bispecific proteins which
CC comprise: (a) a DNA binding domain and (b) a domain which binds a trans-
CC activator (TA), trans-repressor (TR) or their complexes, which are
CC characteristic of a physiological or physiopathological state. The novel
CC chimaeric, bispecific proteins allow expression of a therapeutic protein
CC (e.g. diphtheria or Pseudomonas toxins, thymidine kinase, single chain
CC antibodies) to be regulated in response to particular conditions.
CC Examples include making the protein responsive to the presence of
CC particular pathogenic TA mols (e.g. HIV Tat, papilloma virus E6/E7
CC proteins or Epstein-Barr virus EBNA protein), the therapeutic protein
CC will be expressed in those cells infected by that pathogen. Similarly,
CC where the chimaeric protein responds to a cellular protein typical of a
CC hyperproliferative state (esp. wild-type and mutant p53), expression can
CC be restricted to tumour cells. The sequence presented here is an example
CC of a peptide linker "arm" which connects the DNA binding domain to the TA
CC binding domain
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGSGGGGSGGGGS 15
||||||||||||||||
Db 1 GGGGSGGGGSGGGGS 15

RESULT 8
AAY49219
ID AAY49219 standard; peptide; 15 AA.
XX
AC AAY49219;
XX
DT 07-FEB-2000 (first entry)
XX
DE Sequence of a linking peptide.
XX
KW Monoclonal antibody; MAb; 1A7; GD2; immune response; melanoma;
KW neuroblastoma; glioma; soft tissue carcinoma; small cell carcinoma;
KW tumor-associated antigen.
XX
OS Synthetic.
XX
PN US5977316-A.
XX
PD 02-NOV-1999.
XX
PF 16-JAN-1996; 96US-00591196.
XX
PR 17-JAN-1995; 95US-00372676.
XX
PA (KENT) UNIV KENTUCKY.
XX
PI Foon KA, Chatterjee SK, Chatterjee M;
XX
DR WPI; 1996-354530/35.
XX
PT Monoclonal antibody 1A7 and related polynucleotide(s) and polypeptide(s)
PT - useful to treat or palliate a GD2-associated disease, e.g. melanoma and
PT glioma.
XX
PS Disclosure; Col 24; 74pp; English.
XX
CC The invention provides a monoclonal antibody (MAb) designated 1A7, which
CC elicits an anti-GD2 (tumor-associated antigen) immunological response in
CC humans. MAb 1A7 has defined light and heavy chain variable region
CC sequences. The MAb 1A7 and polypeptides can be used for eliciting an anti
CC -GD2 immune response. The polypeptides can also be used for detecting or
CC purifying anti-GD2 antibody. The products can be used for treating GD2 -
CC associated diseases, e.g. melanoma, neuroblastoma, glioma, soft tissue
CC carcinoma, and small cell carcinoma. They can be used for palliating the
CC disease or for reducing the risk of recurrence
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGSGGGGSGGGGS 15
||||||||||||||||
Db 1 GGGGSGGGGSGGGGS 15

```
RESULT 9
AAW10295
ID   AAW10295 standard; peptide; 15 AA.
XX
AC   AAW10295;
XX
DT   11-SEP-1997  (first entry)
XX
DE   Peptide linker for soluble fused MHC heterodimer:peptide complex.
XX
KW   Soluble; fusion; major histocompatibility complex; MHC; heterodimer;
KW   complex; linker; antigen; binding groove; tolerance; autoantigen;
KW   disease; insulin dependent; diabetes mellitus; IDDM; antagonist; T cell;
KW   anergy; presenting cell.
XX
OS   Synthetic.
XX
PN   WO9640944-A2.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;   96WO-US010102.
XX
PR   07-JUN-1995;   95US-00480002.
PR   07-JUN-1995;   95US-00482133.
PR   07-JUN-1995;   95US-00483241.
PR   27-OCT-1995;   95US-0005964P.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
PA   (ANER-) ANERGEN INC.
XX
PI   Kindsvogel W,  Reich EP,  Gross JA,  Deshpande S,  Sheppard PO;
XX
DR   WPI; 1997-052337/05.
XX
PT   Novel fused major histocompatibility complex:antigenic peptide complex -
PT   useful to induce tolerance to an autoantigen-related disease e.g. insulin
PT   -dependent diabetes mellitus.
XX
PS   Claim 7; Page 137; 142pp; English.
XX
CC   A novel soluble fused major histocompatibility complex (MHC)
CC   heterodimer:peptide complex, comprises DNA encoding 1st and 2nd MHC
CC   domains, linked by DNA encoding a 5-25 residue linker, e.g. the present
CC   peptide, and a DNA encoding an antigenic peptide able to associate with a
CC   peptide binding groove of the MHC molecule, linked in frame to the DNA
CC   encoding the 2nd domain by a DNA encoding a 5-25 residue linker. The
CC   complex can be used to induce immunological tolerance in adults
CC   susceptible to, or suffering from an autoantigen related disease, e.g.
CC   insulin dependent diabetes mellitus (IDDM), by antagonising the binding
CC   of particular T cells and antigen presenting cells, to induce anergy
CC   (immunological non-responsiveness) in the targeted T cell. As the
CC   heterodimers and corresponding antigen are permanently linked into a
CC   single chain, obviating the requirement for complex heterodimer
CC   truncation or formation, the complex eliminates inefficient and non-
CC   specific peptide loading
XX
SQ   Sequence 15 AA;

  Query Match             100.0%;  Score 84;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db          1 GGGGSGGGGSGGGGS 15


RESULT 10
AAW35984
ID   AAW35984 standard; peptide; 15 AA.
XX
AC   AAW35984;
XX
DT   11-MAR-1998  (first entry)
XX
DE   Peptide linker SEQ ID NO:18 from US5683983.
XX
KW   Interleukin 5; IL-5; receptor; inflammatory disease; eosinophil; asthma;
KW   beta-adrenergic agonist; corticosteroid; treatment; mimetic;
KW   primary library.
XX
OS   Synthetic.
XX
PN   US5683983-A.
XX
PD   04-NOV-1997.
XX
PF   07-JUN-1995;   95US-00484083.
XX
PR   07-JUN-1995;   95US-00484083.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Chen M,  Sloan D,  Barrett RW,  England BP,  Schatz PJ;
XX
DR   WPI; 1997-549007/50.
XX
PT   Treatment of disorders mediated by interleukin-5 - by administering
PT   peptide that binds to IL-5 receptor, for treatment of inflammatory
PT   diseases.
XX
PS   Disclosure; Col 41-42; 38pp; English.
XX
CC   A novel method has been developed for treating a disorder mediated by IL-
CC   5 (interleukin-5). The method comprises administering a peptide that
CC   binds to the IL-5 receptor and comprises the following amino acid
CC   sequence, and dimers and oligomers of this: Cys X1 X2 Trp X3 Arg Cys X4
CC   X5 Cys; where X1 = Gly, Ile, Val or Tyr; X2 = Asp or Glu; X3 = Ala or Val
CC   ; X4 = Gln or Pro; and X5 = Ala, Glu, Lys, Met, Asn, Ser or Thr, where
CC   one or more of the CONH linkages may be replaced by a CH2OC(O)NR,
CC   phosphonate, CH2SO2NR, CH2NR, C(O)NR6 or NHCONH linkage, R = H or lower
CC   alkyl and R6 = lower alkyl; the N-terminal group = NRR1, NRCOR, NRCOOR,
CC   NRSO2R, NHCONHR, succinimido or NHCOOCH2Ar; R1 = H or lower alkyl and Ar
CC   = phenyl optionally mono-, di- or tri-substituted by lower alkyl, lower
CC   alkoxy, Cl and Br; the C-terminal group is COR2, R2 = OH, lower alkoxy or
CC   NR3R4, R3 and R4 = H or lower alkyl, or the N atoms of the NR3R4 group
CC   can optionally be part of the amine group of the N-terminus of the
CC   peptide so as to form a cyclic peptide. The present sequence represents a
CC   peptide linker. The peptide causes the production and accumulation of
CC   eosinophils in tissues. It may be used for treating IL-5-mediated
CC   inflammatory disorders, preferably of the respiratory tract, especially
CC   asthma, optionally together with a beta-adrenergic agonist, an
CC   antiinflammatory corticosteroid or ipratropium bromide
XX
SQ   Sequence 15 AA;

  Query Match             100.0%;  Score 84;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.006;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db          1 GGGGSGGGGSGGGGS 15


RESULT 11
AAW87784
ID   AAW87784 standard; peptide; 15 AA.
XX
AC   AAW87784;
XX
DT   11-MAY-1999  (first entry)
XX
DE   Antibody-beta-lactamase fusion protein spacer peptide #1.
XX
```

KW Fusion protein; antibody; light chain; heavy chain; variable region;
KW melanoma-associated antigen; beta-lactamase; cytotoxic agent; prodrug;
KW tumour cell.
XX
OS Synthetic.
XX
PN WO9850432-A1.
XX
PD 12-NOV-1998.
XX
PF 30-APR-1998; 98WO-US008840.
XX
PR 07-MAY-1997; 97US-0045888P.
PR 30-APR-1998; 98US-00070637.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Siemers NO, Yarnold S, Senter PD;
XX
DR WPI; 1999-070092/06.
XX
PT New fusion polypeptide of antibody variable regions and beta-lactamase -
PT are targetted to melanoma-associated antigens and used to generate
PT cytotoxic agents from prodrugs, at tumour cells.
XX
PS Disclosure; Page 6; 50pp; English.
XX
CC The invention relates to a new fusion polypeptide comprising an antibody
CC light and heavy chain variable regions specific for a melanoma-associated
CC antigen (Ag) linked to a beta-lactamase (bL). This peptide represents a
CC spacer peptide used to separate the heavy and light chains of the
CC antibody. The fusion protein is used to deliver cytotoxic agents to
CC tumour cells; it binds to a tumour cell Ag and converts an administered
CC prodrug to the active form
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGSGGGGSGGGGS 15
||||||||||||||||
Db 1 GGGGSGGGGSGGGGS 15

RESULT 12
AAY43414
ID AAY43414 standard; peptide; 15 AA.
XX
AC AAY43414;
XX
DT 20-DEC-1999 (first entry)
XX
DE Peptide SEQ ID NO:13.
XX
KW Angiogenic homology region; AHR; thrombospondin 1; TSP-1; angiostatin;
KW endostatin; anticancer; antiangiogenic; cancer; cardiovascular disease;
KW obesity; osteoarthritis; duodenal ulcer; abnormal neovascularisation;
KW wound healing; arteriosclerosis; ischaemic limb; ischaemic myocardium;
KW diabetes mellitus; blood vessel occlusion.
XX
OS Synthetic.
XX
PN WO9948924-A1.
XX
PD 30-SEP-1999.
XX
PF 23-MAR-1999; 99WO-US006334.
XX
PR 24-MAR-1998; 98US-00046737.
XX
PA (CHIL-) CHILDRENS MEDICAL CENT.
PA (YISS) YISSUM RES & DEV CO.
XX
PI Ben-Sasson SA;
XX
DR WPI; 1999-591075/50.
XX
PT New angiogenic peptide derivatives, used for treating e.g. cancer,
PT cardiovascular diseases, obesity, osteoarthritis, duodenal ulcers,
PT abnormal neovascularization and for wound healing.
XX
PS Disclosure; Page 59; 62pp; English.
XX
CC The present invention specifically describes peptide derivatives
CC comprising an angiogenic homology region (AHR) of endostatin. The peptide
CC derivatives can be used for modulating angiogenesis in humans and
CC animals. The peptides can be used to treat a wide variety of disease
CC conditions, including cancer, cardiovascular diseases (e.g.
CC arteriosclerosis, ischaemic limbs and ischaemic myocardium), obesity,
CC osteoarthritis, duodenal ulcers, abnormal ocular neovascularisation
CC associated e.g. with diabetes mellitus, and to promote wound healing or
CC to stimulate the growth of new blood vessels to bypass, e.g. blood vessel
CC occlusions. The peptide derivatives can also be used for the production
CC of antibodies. The multivalent ligands may enable the administration of
CC lower doses in order to achieve therapeutic efficacy, as compared with a
CC univalent peptide chain. In addition, they can have long in vivo
CC lifetimes and good biodistribution when administered orally or
CC parenterally. The present sequence represents a peptide used in the
CC exemplification of the present invention
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGSGGGGSGGGGS 15
||||||||||||||||
Db 1 GGGGSGGGGSGGGGS 15

RESULT 13
AAY33328
ID AAY33328 standard; protein; 15 AA.
XX
AC AAY33328;
XX
DT 29-NOV-1999 (first entry)
XX
DE E6-sFv peptide linker fragment.
XX
KW Cytotoxic; RNase; ribonuclease; pancreatic; antibody; light chain;
KW heavy chain; cell surface marker; treatment; tumor; viral infection;
KW parasite infection; immune dysfunctional cell; autoimmune disease;
KW contraceptive; cell separation; transplantation; bone marrow ablation;
KW leukemia cell; T-cell; graft-versus-host disease; ss.
XX
OS Synthetic.
XX
PN US5955073-A.
XX
PD 21-SEP-1999.
XX
PF 09-JUL-1997; 97US-00891848.
XX
PR 20-APR-1990; 90US-00510696.
PR 22-OCT-1991; 91US-00779195.
PR 04-FEB-1993; 93US-00014082.
PR 22-SEP-1993; 93US-00125462.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Rybak SM, Newton DL, Nicholls PJ, Youle RJ;
XX

DR WPI; 1999-560488/47.
XX
PT Recombinantly fused pancreatic RNase-targeting proteins useful for
PT treating tumors, infections, immune or autoimmune disorders and as a
PT contraceptive.
XX
PS Example 3; Col 57-58; 47pp; English.
XX
CC This invention describes a novel nucleic acid construct comprising
CC sequences encoding functional pancreatic RNase and a second protein
CC (preferably the light and heavy chains of an antibody) which binds a
CC specific cell surface marker on a target cell and functions as a
CC cytotoxic agent. The products can be used for selectively killing cells
CC expressing a specific surface marker. They can be used for treating
CC tumors or infected cells (e.g. cells infected by viruses (especially
CC latent or chronic virus infections, such as human immunodeficiency virus
CC (HIV)-1, Epstein-Barr virus, herpes viruses (herpes simplex types I and
CC II), hepatitis viruses (B, non-A-non-B, and delta), herpes zoster,
CC cytomegalovirus)) and cells infected with parasites (such as the malaria
CC parasite)). They can also be used for treating immune dysfunctional cells
CC in immune and autoimmune diseases. Additionally, they may be used as
CC contraceptives. Finally they can also be used for cell separation in
CC vitro by selectively killing unwanted types of cells (e.g. in bone
CC marrow) prior to transplantation into a patient undergoing marrow
CC ablation by radiation or for killing leukemia cells or T-cells that would
CC cause graft-versus-host disease. This sequence represents a E6-sFv linker
CC peptide which is used in the method of the invention
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGSGGGGSGGGGS 15
||||||||||||||||
Db 1 GGGGSGGGGSGGGGS 15

RESULT 14
AAY03763
ID AAY03763 standard; peptide; 15 AA.
XX
AC AAY03763;
XX
DT 10-JUN-1999 (first entry)
XX
DE Linker peptide in a single-gene encoding an Ig-like molecule.
XX
KW Signal-chain; immunoglobulin-like molecule; Ig; monoclonal antibody;
KW MAb CC49; human; colorectal; carcinoma; biosensor; gene therapy;
KW linker peptide.
XX
OS Synthetic.
XX
PN US5892019-A.
XX
PD 06-APR-1999.
XX
PF 01-SEP-1994; 94US-00299999.
XX
PR 15-JUL-1987; 87US-00073685.
PR 02-JUL-1990; 90US-00547336.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Kashmiri SVS, Shu L, Schlom J;
XX
DR WPI; 1999-253934/21.
XX
PT Nucleic acid encoding single-chain antibody-like molecule.
XX
PS Claim 28; Col 9; 15pp; English.
XX
CC The invention relates to a single gene encoding a signal-chain
CC immunoglobulin (Ig)-like molecule that has binding affinity for an
CC antigen. The Ig-like molecule comprises (a) the binding part (BP1) of the
CC light chain variable region (Vl); (b) the binding part (BP2) of the heavy
CC chain variable region (Vh); (c) at least one linker connecting BP1 and
CC BP2, and (d) a polypeptide comprising a modified heavy chain constant
CC region; where the Ig-like polypeptide binds to both TAG-72 and LS174T
CC antigens and specifically binds to an epitope also bound by a monoclonal
CC antibody CC49 (ATCC CRL9459); The Ig-like polypeptides have the same
CC specificity, binding properties and cytotoxicity of the parent monoclonal
CC antibody, so are useful for therapy and diagnosis, specifically of human
CC (colorectal) carcinoma and their metastases, also in biosensors, for
CC imaging or for purification. Nucleic acids encoding the Ig-like
CC polypeptides may be used in gene therapy. Use of the nucleic acid
CC eliminates (a) the need to deliver two genes and (ii) problems of
CC inefficient assembly, associated with expression of complete antibodies.
CC The Ig-like polypeptide induce little, if any, anti-murine antibody
CC response, and can be used to transfect cells, e.g. tumour-infiltrating
CC lymphocytes, ex vivo for subsequent delivery to a tumor
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGSGGGGSGGGGS 15
||||||||||||||||
Db 1 GGGGSGGGGSGGGGS 15

RESULT 15
AAY21600
ID AAY21600 standard; peptide; 15 AA.
XX
AC AAY21600;
XX
DT 20-MAR-2003 (revised)
DT 11-AUG-1999 (first entry)
XX
DE EP-919566 peptide Seq ID No: 23.
XX
KW Antimicrobial; oligopeptide; cecropin P1; microbial pathogen; magainin;
KW plant pathogen; food additive; preservative; cosmetic; pharmaceutical.
XX
OS Synthetic.
XX
PN EP919566-A2.
XX
PD 02-JUN-1999.
XX
PF 31-JAN-1992; 98EP-00121780.
XX
PR 01-FEB-1991; 91US-00649784.
PR 31-JAN-1992; 92EP-00101616.
XX
PA (ENIE) ENICHEM SPA.
XX
PI Mapelli C, Dugas De Robertis C, Stahl GF, Bascomb NF;
PI Swerdloff MD, Williams JI, Everett NP;
XX
DR WPI; 1999-304793/26.
XX
PT New oligopeptides containing at least two antimicrobial peptides, useful
PT for protecting plants against microbial pathogens.
XX
PS Disclosure; Fig 1; 67pp; English.
XX
CC The invention relates to antimicrobial peptides including reverse
CC antimicrobial peptides, antimicrobial oligopeptides and other
CC antimicrobial compositions such as cecropin P1. The antimicrobial
CC oligopeptides are active against at least one microbial pathogen, and

CC comprise at least one of a first and one of a second peptide monomer,
CC interconnected directly through a peptide bond via the N and C terminals,
CC or indirectly through a disulfide bond or via bridges. At least one of
CC the first and second monomers confers activity. Oligopeptides connected
CC by bridges do not have the structure of Magainin Pre-pro protein. The
CC antimicrobial peptides are used for providing protection to plants
CC against plant pathogens, thus enhancing crop yields. The peptides are
CC also useful for treatment of human or animal disease, as an additive to
CC foods for preservation, or as a preservative in cosmetics and
CC pharmaceuticals. Unlike prior art antimicrobial peptides Magainins 1 and
CC 2, the new antimicrobial peptides don't have undesirable properties, are
CC not subject to extensive proteolytic degradation, are not phytotoxic to
CC the cell, and have a broader range of activity. (Updated on 20-MAR-2003
CC to correct PF field.) (Updated on 20-MAR-2003 to correct PR field.)
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 84; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGGSGGGGSGGGGS 15
              |||||||||||||||
Db          1 GGGGSGGGGSGGGGS 15
```

Search completed: April 20, 2004, 10:25:02
Job time : 61.6 secs